GENE FAMILIES ASSOCIATED WITH CANCERS

FIELD OF THE INVENTION

The present invention relates to the changes in gene expression in human tissues from cancer patients. The invention specifically relates to human genes which are differentially expressed in cancer tissues of breast, colon, esophagus, kidney, liver, lung, lymph node, ovary, pancreas, prostate, rectum, and/or stomach compared to corresponding normal tissues.

BACKGROUND OF THE INVENTION

In the United States, more than one million new cancer cases are diagnosed and about half million people die of cancer. The causes of cancer are many and varied, and include genetic predisposition, environmental influences, infectious agents and ageing. These transform normal cells into cancerous ones by derailing a wide spectrum of regulatory and downstream effector pathways. Several essential alterations in cell physiology collectively dictate malignant growth: self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of programmed cell death, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan and Weinberg (2000), *Cell* 100:57-70).

To date, researchers have been able to identify many genetic alterations believed to underlie tumor development. These genetic alterations include amplification of oncogenes and mutations that result in the loss of tumor suppressor genes. Oncogenes were initially identified as genes carried by viruses that cause transformation of their target cells. A major class of the viral oncogenes have cellular counterparts that are involved in normal cell functions. The cellular genes are called proto-oncogene, and in certain cases their mutation or aberrant in the cell is associated with tumor formation. The generation of a

oncogene represents a gain-of-function in which a cellular proto-oncogene is inappropriately activated. This can involve a mutational change in the protein, or constitutive activation, over-expression, or failure to turn off expression at the appropriate time. About 100 oncogenes have been identified. Examples of oncogenes include, but are not limited to, ras, fos, myc, abl, and myb (Ponder (2001), *Nature* 411:336-341). Tumor suppressor genes, in their wild-type alleles, express proteins that suppress abnormal cellular proliferation. When the gene coding for a tumor suppressor protein is mutated or deleted, the resulting mutant protein or the complete lack of tumor suppressor protein expression may fail to correctly regulate cellular proliferation, and abnormal proliferation may take place, particularly if there is already existing damage to the cellular regulatory mechanism. A number of well-studied human tumors and tumor cell lines have missing or non-functional tumor suppressor genes. Examples of tumor suppressor genes include, but are not limited to, the retinoblastoma susceptibility gene or RB gene, the p53 gene, the deletion in colon carcinoma (DCC) gene and the neurofibromatosis type 1 (NF-1) tumor suppressor gene (Weinberg (1991), *Science* 254:1138-1146). Loss-of-function or inactivation of tumor suppressor genes may play a central role in the initiation and/or progression of a significant number of human cancers.

The utilization of genome-wide expression profiles to classify tumors, to identify drug targets, to identify diagnostic markers and/or to gain further insights into the consequences of chemotherapeutic treatments could facilitate the design of more efficacious stratagems for treating a variety of cancers. Initial studies utilizing gene expression patterns to identify subtypes of cancer produced rather intriguing results (see Perou *et al.* (1999), *Proc Natl Acad Sci U S A* 96:9212-9217; Golub *et al.* (1999), *Science* 286:531-537; Alizadeh *et al.* (2000), *Nature* 403:503-511; Alon *et al.* (1999), *Proc Natl Acad Sci U S A* 96:6745-6750; and Bittner *et al.* (2000), *Nature* 406:536-540; Perou *et al.* (2000), *Nature* 406:747-752). Molecular classification of B-cell lymphoma by gene expression profiling elucidated clinically distinct diffuse large-B-cell lymphoma subgroups (see Alizadeh *et al.*, supra). In breast cancer, studies utilizing limited numbers

of genes (8,102 genes) have classified tumors into subtypes based on gene expression profiles, and this study indicated a diversity of molecular phenotypes associated with breast tumors (see Perou *et al.*, supra). In addition, the expression profiling has enabled researchers to map tissue-specific expression levels for thousands of genes (Alon *et al.* (1999), *Proc Natl Acad Sci USA* 96:6745-6750; Iyer *et al.* (1999), *Science* 283:83-87; Khan *et al.*(1998), *Cancer Res* 58:5009-5013; Lee *et al.*(1999), *Science* 285:1390-1393; Wang *et al.* (1999), *Gene* 229:101-108; Whitney *et al.* (1999), *Ann Neurol* 46:425-428). Although these studies have demonstrated that expression profiling may be used to produce improvements in diagnosis of human diseases such as cancer, as well as in the development of improved therapeutic strategies, further studies are needed.

Although cancers are diverse and heterogeneous as they are derived from numerous tissues and multiple etiologic factors, it has been suggested that underlying this variability lies a relatively small number of critical events whose convergence is required for the development of any and all cancers (Evan and Vousden (2001), *Nature* 411:342-348). Accordingly, there exists a need for the comprehensive investigation of the changes in global gene expression levels in many different types of cancers to identify critical molecular markers associated with the development and progression of cancer. There remains a need in the art for materials and methods that permit a more accurate diagnosis of cancer. In addition, there remains a need in the art for methods to treat and methods to identify agents that can effectively treat this disease. The present invention meets these and other needs.

SUMMARY OF THE INVENTION

The present invention is based on new genes that are differentially expressed in cancer tissues compared to normal tissues, hereinafter LFG1, LFG2, LFG3, LFG4, LFG5, LFG6, respectively. The invention includes isolated nucleic acid molecules comprising SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 or the complement thereof.

The present invention further includes the nucleic acid molecules operably linked to one or more expression control elements, including vectors comprising the isolated nucleic acid molecules. The invention further includes host cells transformed to contain the nucleic acid molecules of the invention and methods for producing a protein comprising the step of culturing a host cell transformed with a nucleic acid molecule of the invention under conditions in which the protein is expressed.

The invention further provides an isolated polypeptide selected from the group consisting of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, an isolated polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, an isolated polypeptide comprising conservative amino acid substitutions of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 and an isolated polypeptide comprising naturally occurring amino acid sequence variants of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16. Polypeptides of the invention also include polypeptides with an amino acid sequence having at least about 50%, 60%, 70% or 75% amino acid sequence identity with the sequence set forth in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, preferably at least about 80%, more preferably at least about 90-95%, and most preferably at least about 95-98% sequence identity with the sequence set forth in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16.

The present invention further provides methods of identifying other members of the polypeptide family of the invention. Specifically, the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 can be used as a probe, or to generate PCR primers, in methods to identify nucleic acid molecules that encode other members of the LFG1, LFG2, LFG3, LFG4, LFG5 or LFG6 family of proteins.

The invention further provides an isolated antibody or antigen-binding antibody fragment that specifically binds to a polypeptide of the invention, including monoclonal and polyclonal antibodies.

The invention further provides methods of identifying an agent which modulates the expression of a nucleic acid molecule encoding a protein of the invention, comprising: exposing cells which express the nucleic acid molecule to the agent; and determining whether the agent modulates expression of said nucleic acid molecule, thereby identifying an agent which modulates the expression of a nucleic acid molecule encoding the protein.

The invention further provides methods of identifying an agent which modulates the level of or at least one activity of a protein of the invention, comprising: exposing cells which express the protein to the agent; and determining whether the agent modulates the level of or at least one activity of said protein, thereby identifying an agent which modulates the level of or at least one activity of the protein.

The present invention further provides methods of modulating the expression of a nucleic acid molecule encoding a protein of the invention, comprising the step of administering an effective amount of an agent which modulates the expression of a nucleic acid molecule encoding the protein. The invention also provides methods of modulating at least one activity of a protein of the invention, comprising the step of administering an effective amount of an agent which modulates at least one activity of the protein of the invention.

The invention further provides methods of identifying binding partners for a protein of the invention, comprising the steps of exposing said protein to a potential binding partner; and determining if the potential binding partner binds to said protein, thereby identifying binding partners for the protein.

The present invention further provides methods to identify agents that can block or modulate the association of a protein of the invention with a binding partner. Specifically, an agent can be tested for the ability to block, reduce or otherwise modulate the association of a protein of invention with a binding partner by contacting said protein, or a fragment thereof, and a binding partner with a test agent and determining whether the test agent blocks or reduces the binding of the protein of invention to the binding partner.

The present invention further provides methods for reducing or blocking the association of a protein of invention with one or more of its binding partners, comprising the step of administrating an effective amount of an agent which reduces or blocks the binding of said protein to the binding partner. The method can utilize an agent that binds to the protein of invention or to the binding partner.

In accordance with another aspect of the invention, the proteins of the invention can be used as starting points for rational drug design to provide ligands, therapeutic drugs or other types of small chemical molecules. Alternatively, small molecules or other compounds identified by the above-described screening assays may serve as "lead compounds" in rational drug design.

The present invention further relates to a process for treating cancer comprising inserting into a cancerous cell a nucleic acid construct comprising the nucleic acid molecules of the invention operably linked to a promoter or enhancer element such that expression of said nucleic acid molecule causes suppression of said cancer.

The present invention further includes non-human transgenic animals modified to contain the nucleic acid molecules of the invention, or non-human transgenic animals modified to contain the mutated nucleic acid molecules such that expression of the encoded polypeptides of the invention is prevented.

The present invention also includes non-human transgenic animals in which all or a portion of a gene comprising all or a portion of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 has been knocked out or deleted from the genome of the animal.

The invention further provides methods of diagnosing cancers, comprising the steps of acquiring a tissue, blood, urine or other sample from a subject and determining the level of expression of a nucleic acid molecule of the invention or polypeptide of the invention.

The invention further includes compositions comprising a diluent and a polypeptide or protein selected from the group consisting of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, an isolated polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, an isolated polypeptide comprising conservative amino acid substitutions of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, naturally occurring amino acid sequence variants of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 and an isolated polypeptide with an amino acid sequence having at least about 50%, 60%, 70% or 75% amino acid sequence identity with the sequence set forth in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, preferably at least about 80%, more preferably at least about 90-95%, and most preferably at least about 95-98% sequence identity with the sequence set forth in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the relative alignment positions of the two LFG1 clones.

Figure 2 is a hydrophobicity plot of the protein encoded by the open reading frame of LFG1-Clone A (SEQ ID NO: 2). Analysis was performed according to the method of Kyte-Doolittle.

Figure 3 is a hydrophobicity plot of the protein encoded by the open reading frame of LFG1-Clone B (SEQ ID NO: 4). Analysis was performed according to the method of Kyte-Doolittle.

Figure 4 is a hydrophobicity plot of the protein encoded by the open reading frame of LFG2 (SEQ ID NO: 6). Analysis was performed according to the method of Kyte-Doolittle.

Figure 5 is a hydrophobicity plot of the protein encoded by the open reading frame of LFG3 (SEQ ID NO: 8). Analysis was performed according to the method of Kyte-Doolittle.

Figure 6 is a hydrophobicity plot of the protein encoded by the open reading frame of LFG4 (SEQ ID NO: 10). Analysis was performed according to the method of Kyte-Doolittle.

Figure 7 is a hydrophobicity plot of the protein encoded by the open reading frame of ALFG5 (SEQ ID NO: 12). Analysis was performed according to the method of Kyte-Doolittle.

Figure 8 shows the relative alignment positions of the two LFG6 clones.

Figure 9 is a hydrophobicity plot of the protein encoded by the open reading frame of LFG6-#20 (SEQ ID NO: 14). Analysis was performed according to the method of Kyte-Doolittle.

Figure 10 is a hydrophobicity plot of the protein encoded by the open reading frame of LFG6-#46 (SEQ ID NO: 16). Analysis was performed according to the method of Kyte-Doolittle.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

I. General Description

The present invention is based in part on the identification of new gene families that are differentially expressed in cancerous human tissues compared to normal human tissues. These gene families correspond to the human cDNA of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13 and 15.

The genes and proteins of the invention may be used as diagnostic agents or markers to detect cancer or to differentiate carcinoma from normal tissue in a sample. They can also serve as a target for agents that modulate gene expression or activity. For example, agents may be identified that modulate biological processes associated with tumor growth, including the hyperplastic process of cancer.

II. Specific Embodiments

A. The Proteins Associated with Cancer

The present invention provides isolated proteins, allelic variants of the proteins, and conservative amino acid substitutions of the proteins. As used herein, the "protein" or "polypeptide" refers, in part, to a protein that has the human amino acid sequence depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16. The terms also refer to naturally occurring allelic variants and proteins that have a slightly different amino acid sequence than that specifically recited above. Allelic variants, though possessing a slightly different amino acid sequence than those recited above, will still have the same or similar biological functions associated with these proteins.

As used herein, the family of proteins related to the human amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 refers to proteins that have been isolated from organisms in addition to humans. The methods used to identify and isolate other members of the family of proteins related to these proteins are described below.

The proteins of the present invention are preferably in isolated form. As used herein, a protein is said to be isolated when physical, mechanical or chemical methods are employed to remove the protein from cellular constituents that are normally associated with the protein. A skilled artisan can readily employ standard purification methods to obtain an isolated protein.

The proteins of the present invention further include insertion, deletion or conservative amino acid substitution variants of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16. As used herein, a conservative variant refers to alterations in the amino acid sequence that do not adversely affect the biological functions of the protein. A substitution, insertion or deletion is said to adversely affect the protein when the altered sequence prevents or disrupts a biological function associated with the protein. For example, the overall charge, structure or hydrophobic/hydrophilic properties of the protein, in certain instances, may be altered without adversely affecting a biological activity. Accordingly, the amino acid sequence can be altered, for example to render the peptide more hydrophobic or hydrophilic, without adversely affecting the biological activities of the protein.

Ordinarily, the allelic variants, the conservative substitution variants, and the members of the protein family, will have an amino acid sequence having at least about 50%, 60%, 70% or 75% amino acid sequence identity with the sequence set forth in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, more preferably at least about 80%, even more preferably at least about 90-95%, and most preferably at least about 95-98% sequence identity. Identity or homology with respect to such sequences is defined herein as the percentage of amino acid residues in the candidate sequence that are identical with SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent homology, and not considering any conservative substitutions as part of the sequence identity (see section B for the relevant parameters). Fusion proteins, or N-terminal, C-terminal or internal extensions, deletions, or insertions into the peptide sequence shall not be construed as affecting homology.

Thus, the proteins of the present invention include molecules having the amino acid sequence disclosed in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16; fragments thereof having a consecutive sequence of at least about 3, 4, 5, 6, 10, 15, 20, 25, 30, 35 or more amino acid residues of these proteins; amino acid sequence variants wherein one or more amino acid residues has been inserted N- or C-terminal to, or within, the disclosed coding sequence; and amino acid sequence variants of the disclosed sequence, or their fragments

as defined above, that have been substituted by at least one residue. Such fragments, also referred to as peptides or polypeptides, may contain antigenic regions, functional regions of the protein identified as regions of the amino acid sequence which correspond to known protein domains, as well as regions of pronounced hydrophilicity. The regions are all easily identifiable by using commonly available protein sequence analysis software such as MacVector (Oxford Molecular).

Contemplated variants further include those containing predetermined mutations by, *e.g.*, homologous recombination, site-directed or PCR mutagenesis, and the corresponding proteins of other animal species, including but not limited to rabbit, mouse, rat, porcine, bovine, ovine, equine and non-human primate species, and the alleles or other naturally occurring variants of the family of proteins; and derivatives wherein the protein has been covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (for example a detectable moiety such as an enzyme or radioisotope).

The present invention further provides compositions comprising a protein or polypeptide of the invention and a diluent. Suitable diluents can be aqueous or non-aqueous solvents or a combination thereof, and can comprise additional components, for example water-soluble salts or glycerol, that contribute to the stability, solubility, activity, and/or storage of the protein or polypeptide.

As described below, members of the families of proteins can be used: (1) to identify agents which modulate the level of or at least one activity of the protein, (2) to identify binding partners for the protein, (3) as an antigen to raise polyclonal or monoclonal antibodies, (4) as a therapeutic agent or target and (5) as a diagnostic agent or marker of cancer.

B. Nucleic Acid Molecules

The present invention further provides nucleic acid molecules that encode the protein having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 and the related proteins herein described, preferably in isolated form. As used herein, "nucleic acid" is defined as RNA or DNA that encodes a protein or peptide as defined above, is complementary to a nucleic acid sequence encoding such peptides, hybridizes to the nucleic acid of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 and remains stably bound to it under appropriate stringency conditions, encodes a polypeptide sharing at least about 50%, 60%, 70% or 75%, preferably at least about 80%, more preferably at least about 90-95%, and most preferably at least about 95-98% or more identity with the peptide sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 or exhibits at least 50%, 60%, 70% or 75%, preferably at least about 80%, more preferably at least about 90-95%, and most preferably at least about 95-98% or more nucleotide sequence identity over the open reading frames of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15.

The present invention further includes isolated nucleic acid molecules that specifically hybridize to the complement of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15, particularly molecules that specifically hybridize over the open reading frames. Such molecules that specifically hybridize to the complement of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 typically do so under stringent hybridization conditions.

Specifically contemplated are genomic DNA, cDNA, mRNA and antisense molecules, as well as nucleic acids based on alternative backbones or including alternative bases, whether derived from natural sources or synthesized. Such hybridizing or complementary nucleic acids, however, are defined further as being novel and unobvious over any prior art nucleic acid including that which encodes, hybridizes under appropriate stringency conditions, or is complementary to nucleic acid encoding a protein according to the present invention.

Homology or identity at the nucleotide or amino acid sequence level is determined by BLAST (Basic Local Alignment Search Tool) analysis using the algorithm employed by the programs blastp, blastn, blastx, tblastn and tblastx (Altschul *et al.* (1997), *Nucleic*

Acids Res. 25: 3389-3402, and Karlin *et al.* (1990), *Proc. Natl. Acad. Sci. USA* 87: 2264-2268, both fully incorporated by reference) which are tailored for sequence similarity searching. The approach used by the BLAST program is to first consider similar segments, with and without gaps, between a query sequence and a database sequence, then to evaluate the statistical significance of all matches that are identified and finally to summarize only those matches which satisfy a preselected threshold of significance. For a discussion of basic issues in similarity searching of sequence databases, see Altschul *et al.* (1994), *Nat. Genet.* 6: 119-129 which is fully incorporated by reference. The search parameters for histogram, descriptions, alignments, expect (*i.e.*, the statistical significance threshold for reporting matches against database sequences), cutoff, matrix and filter (low complexity) are at the default settings. The default scoring matrix used by blastp, blastx, tblastn, and tblastx is the BLOSUM62 matrix (Henikoff *et al.* (1992), *Proc. Natl. Acad. Sci. USA* 89: 10915-10919, fully incorporated by reference), recommended for query sequences over 85 nucleotides or amino acids in length.

For blastn, the scoring matrix is set by the ratios of M (*i.e.*, the reward score for a pair of matching residues) to N (*i.e.*, the penalty score for mismatching residues), wherein the default values for M and N are 5 and -4, respectively. Four blastn parameters were adjusted as follows: Q=10 (gap creation penalty); R=10 (gap extension penalty); wink=1 (generates word hits at every winkth position along the query); and gapw=16 (sets the window width within which gapped alignments are generated). The equivalent Blastp parameter settings were Q=9; R=2; wink=1; and gapw=32. A Bestfit comparison between sequences, available in the GCG package version 10.0, uses DNA parameters GAP=50 (gap creation penalty) and LEN=3 (gap extension penalty) and the equivalent settings in protein comparisons are GAP=8 and LEN=2.

"Stringent conditions" are those that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50°C, or (2) employ during hybridization a denaturing agent such as formamide, for

example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C. Another example is hybridization in 50% formamide, 5× SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5× Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2× SSC and 0.1% SDS. A skilled artisan can readily determine and vary the stringency conditions appropriately to obtain a clear and detectable hybridization signal. Preferred molecules are those that hybridize under the above conditions to the complement of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 and which encode a functional or full-length protein. Even more preferred hybridizing molecules are those that hybridize under the above conditions to the complement strand of the open reading frame of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15.

As used herein, a nucleic acid molecule is said to be "isolated" when the nucleic acid molecule is substantially separated from contaminant nucleic acid molecules encoding other polypeptides.

The present invention further provides fragments of the disclosed nucleic acid molecules. As used herein, a fragment of a nucleic acid molecule refers to a small portion of the coding or non-coding sequence. The size of the fragment will be determined by the intended use. For example, if the fragment is chosen so as to encode an active portion of the protein, the fragment will need to be large enough to encode the functional region(s) of the protein. For instance, fragments which encode peptides corresponding to predicted antigenic regions may be prepared. If the fragment is to be used as a nucleic acid probe or PCR primer, then the fragment length is chosen so as to obtain a relatively small number of false positives during probing/priming (see the discussion in Section G).

Fragments of the nucleic acid molecules of the present invention (*i.e.*, synthetic oligonucleotides) that are used as probes or specific primers for the polymerase chain reaction (PCR), or to synthesize gene sequences encoding proteins of the invention, can

easily be synthesized by chemical techniques, for example, the phosphoramidite method of Matteucci *et al.*, ((1981) *J. Am. Chem. Soc.* 103: 3185-3191) or using automated synthesis methods. In addition, larger DNA segments can readily be prepared by well known methods, such as synthesis of a group of oligonucleotides that define various modular segments of the gene, followed by ligation of oligonucleotides to build the complete modified gene.

The nucleic acid molecules of the present invention may further be modified so as to contain a detectable label for diagnostic and probe purposes. A variety of such labels are known in the art and can readily be employed with the encoding molecules herein described. Suitable labels include, but are not limited to, biotin, radiolabeled or fluorescently labeled nucleotides and the like. A skilled artisan can readily employ any such label to obtain labeled variants of the nucleic acid molecules of the invention.

C. Isolation of Other Related Nucleic Acid Molecules

As described above, the identification and characterization of the nucleic acid molecule having SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 allows a skilled artisan to isolate nucleic acid molecules that encode other members of the protein family in addition to the sequences herein described. Further, the presently disclosed nucleic acid molecules allow a skilled artisan to isolate nucleic acid molecules that encode other members of the family of proteins in addition to the proteins having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16.

For instance, a skilled artisan can readily use the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 to generate antibody probes to screen expression libraries prepared from appropriate cells. Typically, polyclonal antiserum from mammals such as rabbits immunized with the purified protein (as described below) or monoclonal antibodies can be used to probe a mammalian cDNA or genomic expression library, such as lambda gtll library, to obtain the appropriate coding sequence for other members of the protein family. The cloned cDNA sequence can be expressed as a fusion protein,

expressed directly using its own control sequences, or expressed by constructions using control sequences appropriate to the particular host used for expression of the enzyme.

Alternatively, a portion of the coding sequence herein described can be synthesized and used as a probe to retrieve DNA encoding a member of the protein family from any mammalian organism. Oligomers containing approximately 18-20 nucleotides (encoding about a 6-7 amino acid stretch) are prepared and used to screen genomic DNA or cDNA libraries to obtain hybridization under stringent conditions or conditions of sufficient stringency to eliminate an undue level of false positives.

Additionally, pairs of oligonucleotide primers can be prepared for use in PCR to selectively clone an encoding nucleic acid molecule. A PCR denature/anneal/extend cycle for using such PCR primers is well known in the art and can readily be adapted for use in isolating other encoding nucleic acid molecules.

Nucleic acid molecules encoding other members of the protein family may also be identified in existing genomic or other sequence information using any available computational method, including but not limited to: PSI-BLAST (Altschul *et al.* (1997), *Nucl. Acids Res.* 25: 3389-3402); PHI-BLAST (Zhang *et al.* (1998), *Nucl. Acids Res.* 26: 3986-3990), 3D-PSSM (Kelly *et al.* (2000), *J. Mol. Biol.* 299: 499-520); and other computational analysis methods (Shi *et al.* (1999), *Biochem. Biophys. Res. Commun.* 262: 132-138 and Matsunami *et. al.* (2000), *Nature* 404: 601-604).

D. rDNA molecules Containing a Nucleic Acid Molecule

The present invention further provides recombinant DNA molecules (rDNAs) that contain a coding sequence. As used herein, a rDNA molecule is a DNA molecule that has been subjected to molecular manipulation *in situ*. Methods for generating rDNA molecules are well known in the art, for example, see Sambrook *et al.*, Molecular Cloning- A Laboratory Manual, Third Ed., Cold Spring Harbor Laboratory Press, Cold

Spring Harbor, NY, 2001. In the preferred rDNA molecules, a coding DNA sequence is operably linked to expression control sequences and/or vector sequences.

The choice of vector and/or expression control sequences to which one of the protein family encoding sequences of the present invention is operably linked depends directly, as is well known in the art, on the functional properties desired, *e.g.*, protein expression, and the host cell to be transformed. A vector contemplated by the present invention is at least capable of directing the replication or insertion into the host chromosome, and preferably also expression, of the structural gene included in the rDNA molecule.

Expression control elements that are used for regulating the expression of an operably linked protein encoding sequence are known in the art and include, but are not limited to, inducible promoters, constitutive promoters, secretion signals, and other regulatory elements. Preferably, the inducible promoter is readily controlled, such as being responsive to a nutrient in the host cell's medium.

In one embodiment, the vector containing a coding nucleic acid molecule will include a prokaryotic replicon, *i.e.*, a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extrachromosomally in a prokaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, vectors that include a prokaryotic replicon may also include a gene whose expression confers a detectable marker such as a drug resistance. Typical bacterial drug resistance genes are those that confer resistance to ampicillin, kanamycin, chloramphenicol or tetracycline.

Vectors that include a prokaryotic replicon can further include a prokaryotic or bacteriophage promoter capable of directing the expression (transcription and translation) of the coding gene sequences in a bacterial host cell, such as *E. coli*. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial

hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322 and pBR329 available from BioRad Laboratories, (Richmond, CA), pPL and pKK223 available from Pharmacia (Piscataway, NJ).

Expression vectors compatible with eukaryotic cells, preferably those compatible with vertebrate cells, can also be used to form rDNA molecules that contain a coding sequence. Eukaryotic cell expression vectors, including viral vectors, are well known in the art and are available from several commercial sources. Typically, such vectors are provided containing convenient restriction sites for insertion of the desired DNA segment. Typical of such vectors are pSVL and pKSV-10 (Pharmacia), pBPV-1/pML2d (International Biotechnologies, Inc.), pTDT1 (ATCC, #31255), the vector pCDM8 described herein, and the like eukaryotic expression vectors. Vectors may be modified to include tissue specific promoters if needed.

Eukaryotic cell expression vectors used to construct the rDNA molecules of the present invention may further include a selectable marker that is effective in an eukaryotic cell, preferably a drug resistance selection marker. A preferred drug resistance marker is the gene whose expression results in neomycin resistance, *i.e.*, the neomycin phosphotransferase (*neo*) gene. (Southern *et al.* (1982), *J. Mol. Anal. Genet.* 1:327-341). Alternatively, the selectable marker can be present on a separate plasmid, and the two vectors are introduced by co-transfection of the host cell, and selected by culturing in the appropriate drug for the selectable marker.

E. Host Cells Containing an Exogenously Supplied Coding Nucleic Acid Molecule

The present invention further provides host cells transformed with a nucleic acid molecule that encodes a protein of the present invention. The host cell can be either prokaryotic or eukaryotic. Eukaryotic cells useful for expression of a protein of the invention are not limited, so long as the cell line is compatible with cell culture methods

and compatible with the propagation of the expression vector and expression of the gene product. Preferred eukaryotic host cells include, but are not limited to, yeast, insect and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line. Preferred eukaryotic host cells include Chinese hamster ovary (CHO) cells available from the ATCC as CCL61, NIH Swiss mouse embryo cells (NIH/3T3) available from the ATCC as CRL 1658, baby hamster kidney cells (BHK), and the like eukaryotic tissue culture cell lines.

Any prokaryotic host can be used to express a rDNA molecule encoding a protein of the invention. The preferred prokaryotic host is *E. coli*.

Transformation of appropriate cell hosts with a rDNA molecule of the present invention is accomplished by well known methods that typically depend on the type of vector used and host system employed. With regard to transformation of prokaryotic host cells, electroporation and salt treatment methods are typically employed (see, for example, Cohen *et al.* (1972), *Proc. Natl. Acad. Sci. USA* 69: 2110; and Sambrook *et al.*, *supra*). With regard to transformation of vertebrate cells with vectors containing rDNAs, electroporation, cationic lipid or salt treatment methods are typically employed, see, for example, Graham *et al.* (1973), *Virol.* 52: 456; Wigler *et al.* (1979), *Proc. Natl. Acad. Sci. USA* 76: 1373-1376.

Successfully transformed cells, *i.e.*, cells that contain a rDNA molecule of the present invention, can be identified by well known techniques including the selection for a selectable marker. For example, cells resulting from the introduction of an rDNA of the present invention can be cloned to produce single colonies. Cells from those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA using a method such as that described by Southern, (1975) *J. Mol. Biol.* 98: 503 or Berent *et al.*, (1985) *Biotech.* 3: 208, or the proteins produced from the cell assayed via an immunological method.

F. Production of Recombinant Proteins using a rDNA Molecule

The present invention further provides methods for producing a protein of the invention using nucleic acid molecules herein described. In general terms, the production of a recombinant form of a protein typically involves the following steps:

First, a nucleic acid molecule is obtained that encodes a protein of the invention, such as a nucleic acid molecule comprising, consisting essentially of or consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15, or nucleotides 390-4883 or 390-4880 of SEQ ID NO: 1, or nucleotides 12-4907 or 12-4904 of SEQ ID NO: 3, or nucleotides 424-1911 or 424-1908 of SEQ ID NO: 5, or nucleotides 405-1838 or 405-1835 of SEQ ID NO: 7, or nucleotides 89-1153 or 89-1150 of SEQ ID NO: 9, or nucleotides 223-1572 or 223-1569 of SEQ ID NO: 11, or 418-1395 or 418-1392 of SEQ ID NO: 13, or nucleotides 271-1434 or 271-1431 of SEQ ID NO: 15. If the encoding sequence is uninterrupted by introns, as are these open-reading-frames, it is directly suitable for expression in any host.

The nucleic acid molecule is then preferably placed in operable linkage with suitable control sequences, as described above, to form an expression unit containing the protein open reading frame. The expression unit is used to transform a suitable host and the transformed host is cultured under conditions that allow the production of the recombinant protein. Optionally the recombinant protein is isolated from the medium or from the cells; recovery and purification of the protein may not be necessary in some instances where some impurities may be tolerated.

Each of the foregoing steps can be done in a variety of ways. For example, the desired coding sequences may be obtained from genomic fragments and used directly in appropriate hosts. The construction of expression vectors that are operable in a variety of hosts is accomplished using appropriate replicons and control sequences, as set forth above. The control sequences, expression vectors, and transformation methods are dependent on the type of host cell used to express the gene and were discussed in detail earlier. Suitable restriction sites can, if not normally available, be added to the ends of the

coding sequence so as to provide an excisable gene to insert into these vectors. A skilled artisan can readily adapt any host/expression system known in the art for use with the nucleic acid molecules of the invention to produce recombinant protein.

G. Methods to Identify Agents that Modulate the Expression of a Nucleic Acid Encoding the Genes Associated with Cancer

Another embodiment of the present invention provides methods for identifying agents that modulate the expression of a nucleic acid encoding a protein of the invention such as a protein having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16. Such assays may utilize any available means of monitoring for changes in the expression level of the nucleic acids of the invention. As used herein, an agent is said to modulate the expression of a nucleic acid of the invention if it is capable of up- or down-regulating expression of the nucleic acid in a cell.

In one assay format, cell lines that contain reporter gene fusions between nucleotides from within the open reading frame defined by nucleotides 390-4883 of SEQ ID NO: 1, nucleotides 12-4907 of SEQ ID NO: 3, nucleotides 424-1911 of SEQ ID NO: 5, nucleotides 405-1838 of SEQ ID NO: 7, nucleotides 89-1153 of SEQ ID NO: 9, nucleotides 223-1572 of SEQ ID NO: 11, nucleotides 418-1395 of SEQ ID NO: 13, nucleotides 271-1434 of SEQ ID NO: 15, and/or the 5' and/or 3' regulatory elements and any assayable fusion partner may be prepared. Numerous assayable fusion partners are known and readily available including the firefly luciferase gene and the gene encoding chloramphenicol acetyltransferase (Alam *et al.* (1990), *Anal. Biochem.* 188: 245-254). Cell lines containing the reporter gene fusions are then exposed to the agent to be tested under appropriate conditions and time. Differential expression of the reporter gene between samples exposed to the agent and control samples identifies agents which modulate the expression of a nucleic acid of the invention.

Additional assay formats may be used to monitor the ability of the agent to modulate the expression of a nucleic acid encoding a protein of the invention, such as the protein having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16. For instance, mRNA expression may be monitored directly by hybridization to the nucleic acids of the invention. Cell lines are exposed to the agent to be tested under appropriate conditions and time and total RNA or mRNA is isolated by standard procedures such those disclosed in Sambrook *et al.*, Molecular Cloning - A Laboratory Manual, Third Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.

The preferred cells will be those derived from human tissue, for instance, biopsy tissue or cultured cells from patients with cancer. Cell lines such as ATCC breast ductal carcinoma cell lines (Catalogue Nos. CRL-2320, CRL-2338, and CRL-7345), ATCC colorectal adenocarcinoma cell lines (Catalogue Nos. CCL-222, CCL-224, CCL-225, CCL-234, CRL-7159, and CRL-7184), ATCC kidney clear cell carcinoma cell lines (Catalogue Nos. HTB-46 and HTB-47), ATCC renal cell adenocarcinoma cell lines (Catalogue Nos. CRL-1611, CRL-1932 and CRL-1933), ATCC liver hepatocellular carcinoma cell lines (Catalogue Nos. CRL-2233, CRL-2234, and HB-8065), ATCC lung adenocarcinoma cell lines (Catalogue Nos. CRL-5944, CRL-7380, and CRL-5907), ATCC lymphoma cell lines (Catalogue Nos. CRL-7936, CRL-7264, and CRL-7507), ATCC ovary adenocarcinoma cell lines (Catalogue Nos. HTB-161, HTB-75, and HTB-76), ATCC pancreas adenocarcinoma cell lines (Catalogue Nos. CRL-1687, CRL-2119, and HTP-79), prostate adenocarinoma cell lines (Catalogue Nos. CRL-1435, CRL-2422, and CRL-2220), and ATCC gastric adenocarcinoma cell lines (Catalogue Nos. CRL-1739, CRL-1863, and CRL-1864) may be used. Alternatively, other available cells or cell lines may be used.

Probes to detect differences in RNA expression levels between cells exposed to the agent and control cells may be prepared from the nucleic acids of the invention. It is preferable, but not necessary, to design probes which hybridize only with target nucleic acids under conditions of high stringency. Only highly complementary nucleic acid

hybrids form under conditions of high stringency. Accordingly, the stringency of the assay conditions determines the amount of complementarity which should exist between two nucleic acid strands in order to form a hybrid. Stringency should be chosen to maximize the difference in stability between the probe:target hybrid and probe:non-target hybrids.

Probes may be designed from the nucleic acids of the invention through methods known in the art. For instance, the G+C content of the probe and the probe length can affect probe binding to its target sequence. Methods to optimize probe specificity are commonly available in Sambrook *et al.*, *supra*, or Ausubel *et al.*, Short Protocols in Molecular Biology, Fourth Ed., John Wiley & Sons, Inc., New York, 1999.

Hybridization conditions are modified using known methods, such as those described by Sambrook *et al.* and Ausubel *et al.* as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA RNA can be accomplished in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support and the solid support exposed to at least one probe comprising at least one, or part of one of the sequences of the invention under conditions in which the probe will specifically hybridize. Alternatively, nucleic acid fragments comprising at least one, or part of one of the sequences of the invention can be affixed to a solid support, such as a silicon chip, porous glass wafer or membrane. The solid support can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences will specifically hybridize. Such solid supports and hybridization methods are widely available, for example, those disclosed by Beattie, (1995) WO 95/11755. By examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up- or down-regulate the expression of a nucleic acid encoding the protein having the sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 are identified.

Hybridization for qualitative and quantitative analysis of mRNAs may also be carried out by using a RNase Protection Assay (*i.e.*, RPA, see Ma *et al.* (1996), *Methods* 10: 273-238). Briefly, an expression vehicle comprising cDNA encoding the gene product and a phage specific DNA dependent RNA polymerase promoter (*e.g.*, T7, T3 or SP6 RNA polymerase) is linearized at the 3' end of the cDNA molecule, downstream from the phage promoter, wherein such a linearized molecule is subsequently used as a template for synthesis of a labeled antisense transcript of the cDNA by *in vitro* transcription. The labeled transcript is then hybridized to a mixture of isolated RNA (*i.e.*, total or fractionated mRNA) by incubation at 45°C overnight in a buffer comprising 80% formamide, 40 mM Pipes, pH 6.4, 0.4 M NaCl and 1 mM EDTA. The resulting hybrids are then digested in a buffer comprising 40 μg/ml ribonuclease A and 2 μg/ml ribonuclease. After deactivation and extraction of extraneous proteins, the samples are loaded onto urea/polyacrylamide gels for analysis.

In another assay, to identify agents which affect the expression of the instant gene products, cells or cell lines are first identified which express the gene products of the invention physiologically. Cells and/or cell lines so identified would be expected to comprise the necessary cellular machinery such that the fidelity of modulation of the transcriptional apparatus is maintained with regard to exogenous contact of agent with appropriate surface transduction mechanisms and/or the cytosolic cascades. Further, such cells or cell lines would be transduced or transfected with an expression vehicle (*e.g.*, a plasmid or viral vector) construct comprising an operable non-translated 5' promoter-containing end of the structural gene encoding the instant gene products fused to one or more antigenic fragments, which are peculiar to the instant gene products, wherein said fragments are under the transcriptional control of said promoter and are expressed as polypeptides whose molecular weight can be distinguished from the naturally occurring polypeptides or may further comprise an immunologically distinct tag or other detectable marker. Such a process is well known in the art (see Sambrook *et al.*, *supra*).

Cells or cell lines transduced or transfected as outlined above are then contacted with agents under appropriate conditions. For example, the agent in a pharmaceutically acceptable excipient is contacted with cells in an aqueous physiological buffer such as phosphate buffered saline (PBS) at physiological pH, Eagles balanced salt solution (BSS) at physiological pH, PBS or BSS comprising serum or conditioned media comprising PBS or BSS and/or serum incubated at 37°C. Said conditions may be modulated as deemed necessary by one of skill in the art. Subsequent to contacting the cells with the agent, said cells will be disrupted and the polypeptides of the lysate are fractionated such that a polypeptide fraction is pooled and contacted with an antibody to be further processed by immunological assay (*e.g.*, ELISA, immunoprecipitation or Western blot). The pool of proteins isolated from the "agent-contacted" sample will be compared with a control sample where only the excipient is contacted with the cells and an increase or decrease in the immunologically generated signal from the "agent-contacted" sample compared to the control will be used to distinguish the effectiveness of the agent.

H. Methods to Identify Agents that Modulate the Level or at Least One Activity of the Cancer Associated Proteins

Another embodiment of the present invention provides methods for identifying agents that modulate the level or at least one activity of a protein of the invention such as the protein having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16. Such methods or assays may utilize any means of monitoring or detecting the desired activity and are particularly useful for identifying agents that treat cancer.

In one format, the relative amounts of a protein of the invention between a cell population that has been exposed to the agent to be tested compared to an un-exposed control cell population may be assayed. In this format, probes such as specific antibodies are used to monitor the differential expression of the protein in the different cell populations. Cell lines or populations are exposed to the agent to be tested under appropriate conditions and time. Cellular lysates may be prepared from the exposed cell

line or population and a control, unexposed cell line or population. The cellular lysates are then analyzed with the probe.

Antibody probes are prepared by immunizing suitable mammalian hosts in appropriate immunization protocols using the peptides, polypeptides or proteins of the invention if they are of sufficient length, or, if desired, or if required to enhance immunogenicity, conjugated to suitable carriers. Methods for preparing immunogenic conjugates with carriers such as BSA, KLH, or other carrier proteins are well known in the art. In some circumstances, direct conjugation using, for example, carbodiimide reagents may be effective; in other instances linking reagents such as those supplied by Pierce Chemical Co. (Rockford, IL), may be desirable to provide accessibility to the hapten. The hapten peptides can be extended at either the amino or carboxy terminus with a cysteine residue or interspersed with cysteine residues, for example, to facilitate linking to a carrier. Administration of the immunogens is conducted generally by injection over a suitable time period and with use of suitable adjuvants, as is generally understood in the art. During the immunization schedule, titers of antibodies are taken to determine adequacy of antibody formation.

While the polyclonal antisera produced in this way may be satisfactory for some applications, for pharmaceutical compositions, use of monoclonal preparations is preferred. Immortalized cell lines which secrete the desired monoclonal antibodies may be prepared using the standard method of Kohler and Milstein ((1975) *Nature* 256: 495-497) or modifications which effect immortalization of lymphocytes or spleen cells, as is generally known. The immortalized cell lines secreting the desired antibodies are screened by immunoassay in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production in ascites fluid.

The desired monoclonal antibodies are then recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonal antibodies or the polyclonal

antisera which contain the immunologically significant (antigen-binding) portion can be used as antagonists, as well as the intact antibodies. Use of immunologically reactive (antigen-binding) antibody fragments, such as the Fab, Fab', or $F(ab')_2$ fragments is often preferable, especially in a therapeutic context, as these fragments are generally less immunogenic than the whole immunoglobulin.

The antibodies or antigen-binding fragments may also be produced, using current technology, by recombinant means. Antibody regions that bind specifically to the desired regions of the protein can also be produced in the context of chimeras with multiple species origin, such as humanized antibodies.

Agents that are assayed in the above method can be randomly selected or rationally selected or designed. As used herein, an agent is said to be randomly selected when the agent is chosen randomly without considering the specific sequences involved in the association of a protein of the invention alone or with its associated substrates, binding partners, etc. An example of randomly selected agents is the use a chemical library or a peptide combinatorial library, or a growth broth of an organism.

As used herein, an agent is said to be rationally selected or designed when the agent is chosen on a nonrandom basis which takes into account the sequence of the target site and/or its conformation in connection with the agent's action. Agents can be rationally selected or rationally designed by utilizing the peptide sequences that make up these sites. For example, a rationally selected peptide agent can be a peptide whose amino acid sequence is identical to or a derivative of any functional consensus site.

The agents of the present invention can be, as examples, peptides, small molecules, vitamin derivatives, as well as carbohydrates. Dominant negative proteins, DNAs encoding these proteins, antibodies to these proteins, peptide fragments of these proteins or mimics of these proteins may be introduced into cells to affect function. "Mimic" used herein refers to the modification of a region or several regions of a peptide molecule to provide a structure chemically different from the parent peptide but topographically and

functionally similar to the parent peptide (see Grant in: Molecular Biology and Biotechnology, Meyers, ed., pp. 659-664, VCH Publishers, Inc., New York, 1995). A skilled artisan can readily recognize that there is no limit as to the structural nature of the agents of the present invention.

The peptide agents of the invention can be prepared using standard solid phase (or solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides may be synthesized using commercially available oligonucleotide synthesis instrumentation and produced recombinantly using standard recombinant production systems. The production using solid phase peptide synthesis is necessitated if non-gene-encoded amino acids are to be included.

Another class of agents of the present invention are antibodies immunoreactive with critical positions of proteins of the invention, e.g., cytoplasmic domain, spacer domain, α-helical coiled-coil domain, or the receptor domain, as described herein. Antibody agents are obtained by immunization of suitable mammalian subjects with peptides, containing as antigenic regions, those portions of the protein intended to be targeted by the antibodies.

I. Uses for Agents that Modulate the Expression or at Least one Activity of the Proteins Associated with Cancer

As provided in the Examples, the proteins and nucleic acids of the invention, such as the proteins having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, are differentially expressed in cancerous tissue. Agents that up- or down- regulate or modulate the expression of the protein or at least one activity of the protein, such as agonists or antagonists, may be used to modulate biological and pathologic processes associated with the protein's function and activity. This includes agents identified employing homologues and analogues of the present invention.

As used herein, a subject can be any mammal, so long as the mammal is in need of modulation of a pathological or biological process mediated by a protein of the invention. The term "mammal" is defined as an individual belonging to the class Mammalia. The invention is particularly useful in the treatment of human subjects.

Pathological processes refer to a category of biological processes which produce a deleterious effect. For example, expression of a protein of the invention may be associated with cell growth or hyperplasia. As used herein, an agent is said to modulate a pathological process when the agent reduces the degree or severity of the process. For instance, cancer may be prevented or disease progression modulated by the administration of agents which up- or down-regulate or modulate in some way the expression or at least one activity of a protein of the invention.

The agents of the present invention can be provided alone, or in combination with other agents that modulate a particular pathological process. For example, an agent of the present invention can be administered in combination with other known drugs. As used herein, two agents are said to be administered in combination when the two agents are administered simultaneously or are administered independently in a fashion such that the agents will act at the same time.

The agents of the present invention can be administered via parenteral, subcutaneous, intravenous, intramuscular, intraperitoneal, transdermal, or buccal routes. Alternatively, or concurrently, administration may be by the oral route. The dosage administered will be dependent upon the age, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired.

The present invention further provides compositions containing one or more agents which modulate expression or at least one activity of a protein of the invention. While individual needs vary, determination of optimal ranges of effective amounts of each component is within the skill of the art. Typical dosages comprise 0.1 to 100 μg/kg body

wt. The preferred dosages comprise 0.1 to 10 μg/kg body wt. The most preferred dosages comprise 0.1 to 1 μg/kg body wt.

In addition to the pharmacologically active agent, the compositions of the present invention may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically for delivery to the site of action. Suitable formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions may be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol, and/or dextran. Optionally, the suspension may also contain stabilizers. Liposomes can also be used to encapsulate the agent for delivery into the cell.

The pharmaceutical formulation for systemic administration according to the invention may be formulated for enteral, parenteral or topical administration. Indeed, all three types of formulations may be used simultaneously to achieve systemic administration of the active ingredient.

Suitable formulations for oral administration include hard or soft gelatin capsules, pills, tablets, including coated tablets, elixirs, suspensions, syrups or inhalations and controlled release forms thereof.

In practicing the methods of this invention, the compounds of this invention may be used alone or in combination, or in combination with other therapeutic or diagnostic agents. In certain preferred embodiments, the compounds of this invention may be coadministered along with other compounds typically prescribed for these conditions according to generally accepted medical practice. The compounds of this invention can be

utilized *in vivo*, ordinarily in mammals, such as humans, sheep, horses, cattle, pigs, dogs, cats, rats and mice, or *in vitro*.

J. Methods to Identify Binding Partners

Another embodiment of the present invention provides methods for isolating and identifying binding partners of proteins of the invention. In general, a protein of the invention is mixed with a potential binding partner or an extract or fraction of a cell under conditions that allow the association of potential binding partners with the protein of the invention. After mixing, peptides, polypeptides, proteins or other molecules that have become associated with a protein of the invention are separated from the mixture. The binding partner that bound to the protein of the invention can then be removed and further analyzed. To identify and isolate a binding partner, the entire protein, for instance a protein comprising the entire amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 can be used. Alternatively, a fragment of the protein can be used.

As used herein, a cellular extract refers to a preparation or fraction which is made from a lysed or disrupted cell. The preferred source of cellular extracts will be cells derived from human tumors or transformed cells, for instance, biopsy tissue or tissue culture cells from carcinomas. Alternatively, cellular extracts may be prepared from normal tissue or available cell lines.

A variety of methods can be used to obtain an extract of a cell. Cells can be disrupted using either physical or chemical disruption methods. Examples of physical disruption methods include, but are not limited to, sonication and mechanical shearing. Examples of chemical lysis methods include, but are not limited to, detergent lysis and enzyme lysis. A skilled artisan can readily adapt methods for preparing cellular extracts in order to obtain extracts for use in the present methods.

Once an extract of a cell is prepared, the extract is mixed with the protein of the invention under conditions in which association of the protein with the binding partner can

occur. A variety of conditions can be used, the most preferred being conditions that closely resemble conditions found in the cytoplasm of a human cell. Features such as osmolarity, pH, temperature, and the concentration of cellular extract used, can be varied to optimize the association of the protein with the binding partner.

After mixing under appropriate conditions, the bound complex is separated from the mixture. A variety of techniques can be utilized to separate the mixture. For example, antibodies specific to a protein of the invention can be used to immunoprecipitate the binding partner complex. Alternatively, standard chemical separation techniques such as chromatography and density/sediment centrifugation can be used.

After removal of non-associated cellular constituents found in the extract, the binding partner can be dissociated from the complex using conventional methods. For example, dissociation can be accomplished by altering the salt concentration or pH of the mixture.

To aid in separating associated binding partner pairs from the mixed extract, the protein of the invention can be immobilized on a solid support. For example, the protein can be attached to a nitrocellulose matrix or acrylic beads. Attachment of the protein to a solid support aids in separating peptide/binding partner pairs from other constituents found in the extract. The identified binding partners can be either a single protein or a complex made up of two or more proteins. Alternatively, binding partners may be identified using a Far-Western assay according to the procedures of Takayama *et al.* (1997), *Methods Mol. Biol.* 69: 171-184 or Sauder *et al.* (1996), *J. Gen. Virol.* 77: 991-996 or identified through the use of epitope tagged proteins or GST fusion proteins.

Alternatively, the nucleic acid molecules of the invention can be used in a yeast two-hybrid system or other *in vivo* protein-protein detection system. The yeast two-hybrid system has been used to identify other protein partner pairs and can readily be adapted to employ the nucleic acid molecules herein described.

K. Use of the Binding Partners of the Cancer Associated Proteins

Once isolated, the binding partners of the proteins of the invention, and homologues and analogues thereof, obtained using the above described methods can be used for a variety of purposes. The binding partners can be used to generate antibodies that bind to the binding partner using techniques known in the art. Antibodies that bind the binding partner can be used to assay the activity of the protein of the invention, as a therapeutic agent to modulate a biological or pathological process mediated by the protein of the invention, or to purify the binding partner. These uses are described in detail below.

L. Methods to Identify Agents that Block the Associations between the Binding Partners and the Cancer Associated Proteins

Another embodiment of the present invention provides methods for identifying agents that reduce or block the association of a protein of the invention with a binding partner. Specifically, a protein of the invention is mixed with a binding partner in the presence and absence of an agent to be tested. After mixing under conditions that allow association of the proteins, the two mixtures are analyzed and compared to determine if the agent reduced or blocked the association of the protein of the invention with the binding partner. Agents that block or reduce the association of the protein of the invention with the binding partner will be identified as decreasing the amount of association present in the sample containing the tested agent.

As used herein, an agent is said to reduce or block the association between a protein of the invention and a binding partner when the presence of the agent decreases the extent to which or prevents the binding partner from becoming associated with the protein of the invention. One class of agents will reduce or block the association by binding to the binding partner while another class of agents will reduce or block the association by binding to the protein of the invention.

The binding partner used in the above assay can either be an isolated and fully characterized protein or can be a partially characterized protein that binds to the protein of the invention or a binding partner that has been identified as being present in a cellular extract. It will be apparent to one of ordinary skill in the art that so long as the binding partner has been characterized by an identifiable property, e.g., molecular weight, the present assay can be used.

Agents that are assayed in the above method can be randomly selected or rationally selected or designed. As used herein, an agent is said to be randomly selected when the agent is chosen randomly without considering the specific sequences involved in the association of the protein of the invention with the binding partner. An example of randomly selected agents is the use of a chemical library or a peptide combinatorial library, or a growth broth of an organism.

As used herein, an agent is said to be rationally selected or designed when the agent is chosen on a nonrandom basis which takes into account the sequence of the target site and/or its conformation in connection with the agent's action. Agents can be rationally selected or rationally designed by utilizing the peptide sequences that make up the contact sites of the binding partner with the protein of the invention. For example, a rationally selected peptide agent can be a peptide whose amino acid sequence is identical to the contact site of the protein of the invention on the binding partner. Such an agent will reduce or block the association of the protein of the invention with the binding partner by binding to the binding partner.

The agents of the present invention can be, as examples, peptides, small molecules, vitamin derivatives, as well as carbohydrates. A skilled artisan can readily recognize that there is no limit as to the structural nature of the agents of the present invention.

One class of agents of the present invention are peptide agents whose amino acid sequences are chosen based on the amino acid sequence of the protein of the invention. The peptide agents of the invention can be prepared using standard solid phase (or

solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides may be synthesized using commercially available oligonucleotide synthesis instrumentation and produced recombinantly using standard recombinant production systems. The production using solid phase peptide synthesis is necessitated if non-gene encoded amino acids are to be included.

Another class of agents of the present invention are antibodies immunoreactive with critical positions of the protein of the invention or the binding partner. As described above, antibodies are obtained by immunization of suitable mammalian subjects with peptides, containing as antigenic regions, those portions of the protein of the invention or the binding partner, intended to be targeted by the antibodies. Critical regions include the contact sites involved in the association of the protein of the invention with the binding partner.

As discussed below, the important minimal sequence of residues involved in activity of the protein of the invention define a functional linear domain that can be effectively used as a bait for two hybrid screening and identification of potential associated molecules. Use of such fragments will significantly increase the specificity of the screening as opposed to using the full-length molecule and is therefore preferred. Similarly, this linear sequence can be also used as an affinity matrix also to isolate binding proteins using a biochemical affinity purification strategy.

M. Uses for Agents that Block the Associations between the Binding Partners and the Cancer Associated Proteins

As provided in the Examples, the proteins and nucleic acids of the invention, such as the proteins having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, are differentially expressed in cancerous tissue. Agents that reduce or block the interactions of a protein of the invention, including those identified employing

homologues and analogues of the protein, with a binding partner may be used to modulate biological and pathologic processes associated with the protein's function and activity.

As used herein, a subject can be any mammal, so long as the mammal is in need of modulation of a pathological or biological process mediated by a protein of the invention. The term "mammal" is meant an individual belonging to the class Mammalia. The invention is particularly useful in the treatment of human subjects.

Pathological processes refer to a category of biological processes which produce a deleterious effect. For example, expression of a protein of the invention may be associated with cell growth or hyperplasia. As used herein, an agent is said to modulate a pathological process when the agent reduces the degree or severity of the process. For instance, cancer may be prevented or disease progression modulated by the administration of agents that reduce or block the interactions of a protein of the invention with a binding partner.

The agents of the present invention can be administered via parenteral, subcutaneous, intravenous, intramuscular, intraperitoneal, transdermal, or buccal routes. Alternatively, or concurrently, administration may be by the oral route. The dosage administered will be dependent upon the age, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired.

The present invention further provides compositions containing one or more agents that block association of a protein of the invention with a binding partner. While individual needs vary, determination of optimal ranges of effective amounts of each component is within the skill of the art. Typical dosages comprise 0.1 to 100 μg/kg body wt. The preferred dosages comprise 0.1 to 10 μg/kg body wt. The most preferred dosages comprise 0.1 to 1 μg/kg body wt.

In addition to the pharmacologically active agent, the compositions of the present invention may contain suitable pharmaceutically acceptable carriers comprising excipients

and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically for delivery to the site of action. Suitable formulations for parenteral administration include aqueous solutions of the active compounds in water soluble form, for example, water soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions may be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol, and/or dextran. Optionally, the suspension may also contain stabilizers. Liposomes can also be used to encapsulate the agent for delivery into the cell.

The pharmaceutical formulation for systemic administration according to the invention may be formulated for enteral, parenteral or topical administration. Indeed, all three types of formulations may be used simultaneously to achieve systemic administration of the active ingredient.

Suitable formulations for oral administration include hard or soft gelatin capsules, pills, tablets, including coated tablets, elixirs, suspensions, syrups or inhalations and controlled release forms thereof.

In practicing the methods of this invention, the compounds of this invention may be used alone or in combination, or in combination with other therapeutic or diagnostic agents. In certain preferred embodiments, the compounds of this invention may be coadministered along with other compounds typically prescribed for these conditions according to generally accepted medical practice. The compounds of this invention can be utilized *in vivo*, ordinarily in mammals, such as humans, sheep, horses, cattle, pigs, dogs, cats, rats and mice, or *in vitro*.

N. Rational Drug Design and Combinatorial Chemistry

The present invention further encompasses rational drug design and combinatorial chemistry. Those of skill will recognize appropriate methods to utilize and exploit aspects of the present invention in identifying compounds which can be developed for cancer treatment. Rational drug design involving polypeptides requires identifying and defining a first peptide with which the designed drug is to interact, and using the first target peptide to define the requirements for a second peptide. With such requirements defined, one can find or prepare an appropriate peptide or non-peptide that meets all or substantially all of the defined requirements. Thus, one goal of rational drug design is to produce structural or functional analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, null compounds) in order to fashion drugs that are, for example, more or less potent forms of the ligand. (See, e.g., Hodgson (1991), *Bio. Technology* 9:19-21). Combinatorial chemistry is the science of synthesizing and testing compounds for bioactivity en masse, instead of one by one, the aim being to discover drugs and materials more quickly and inexpensively than was formerly possible. Rational drug design and combinatorial chemistry have become more intimately related in recent years due to the development of approaches in computer-aided protein modeling and drug discovery. (See e.g., US Pat. No. 4,908,773; 5,884,230; 5,873,052; 5,331,573; and 5,888,738).

The use of molecular modeling as a tool for rational drug design and combinatorial chemistry has dramatically increased due to the advent of computer graphics. Not only is it possible to view molecules on computer screens in three dimensions but it is also possible to examine the interactions of macromolecules such as enzymes and receptors and rationally designed derivative molecules to test. (See Boorman (1992), *Chem. Eng. News* 70:18-26). A vast amount of user-friendly software and hardware is now available and virtually all pharmaceutical companies have computer modeling groups devoted to rational drug design. Molecular Simulations Inc. (www.msi.com), for example, sells several sophisticated programs that allow a user to start from an amino acid sequence, build a two or three-dimensional model of the protein or polypeptide, compare it to other

two and three-dimensional models, and analyze the interactions of compounds, drugs, and peptides with a three dimensional model in real time. Accordingly, in some embodiments of the invention, software is used to compare regions of the invention protein and molecules that interact therewith (collectively referred to as "binding partners" --e.g., anti-protein antibodies), and fragments or derivatives of these molecules with other molecules, such as peptides, peptidomimetics, and chemicals, so that therapeutic interactions can be predicted and designed. (See Schneider (1998), *Genetic Engineering News* December: page 20; Tempczyk *et al.* (1997), Molecular Simulations Inc. Solutions April; and Butenhof (1998), Molecular Simulations Inc. Case Notes (August 1998) for a discussion of molecular modeling).

O. Gene Therapy

In another embodiment, genetic therapy can be used as a means for modulating biological and pathologic processes associated with the protein's function and activity. This comprises inserting into a cancerous cell a gene construct encoding a protein comprising all or at least a portion of the sequences of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, or alternatively a gene construct comprising all or a portion of the non-coding region of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15, operably linked to a promoter or enhancer element such that expression of said protein causes suppression of said cancer and wherein said promoter or enhancer element is a promoter or enhancer element modulating said gene construct.

In the constructs described, expression of said protein can be directed from any suitable promoter (e.g., the human cytomegalovirus (CMV), simian virus 40 (SV40), or metallothionein promoters), and regulated by any appropriate mammalian regulatory element. For example, if desired, enhancers known to preferentially direct gene expression in neural cells, T cells, or B cells may be used to direct the expression. The enhancers used could include, without limitation, those that are characterized as tissue or cell specific in their expression. Alternatively, if a genomic clone of LFG1, LFG2, LFG3, LFG4, LFG5

or LFG6 is used as a therapeutic construct (for example, following its isolation by hybridization with the nucleic acid molecule of the invention described above), regulation may be mediated by the cognate regulatory sequences or, if desired, by regulatory sequences derived from a heterologous source, including any of the promoters or regulatory elements described above.

Insertion of the construct into a cancerous cell is accomplished *in vivo*, for example using a viral or plasmid vector. Such methods can also be applied to *in vitro* uses. Thus, the methods of the present invention are readily applicable to different forms of gene therapy, either where cells are genetically modified *ex vivo* and then administered to a host or where the gene modification is conducted *in vivo* using any of a number of suitable methods involving vectors especially suitable to such therapies.

Retroviral vectors, adenoviral vectors, adeno-associated viral vectors, or other viral vectors with the appropriate tropism for cells likely to be involved in cancer (for example, epithelial cells) may be used as a gene transfer delivery system for a therapeutic gene construct. Numerous vectors useful for this purpose are generally known (Cozzi PJ, et al., (2002) *Prostate*, 53(2):95-100; Bitzer M, Lauer U., (2002) *Dtsch Med Wochenschr.* 127(31-32):1623-1624; Mezzina and Danos (2002), *Trends Genet.* 8:241-256; Loser *et al.* (2002) *Curr. Gene Ther.* 2:161-171; Pfeifer and Verma (2001), *Annu. Rev. Genomics Hum. Genet.* 2:177-211). Retroviral vectors are particularly well developed and have been used in clinical settings (Anderson *et al.* (1995), U.S. Patent No. 5,399,346). Non-viral approaches may also be employed for the introduction of therapeutic DNA into cells otherwise predicted to undergo cancer (Jeschke *et al.* (20002) *Curr. Gene Ther.* 1:267-278; Wu *et al.* (1988), *J. Biol. Chem.* 263:14621-14624; Wu *et al.* (1989), *J. Biol. Chem.* 264:16985-16987). For example, a gene may be introduced into a neuron or a T cell by lipofection, asialorosonucoid polylysine conjugation, or, less preferably, microinjection under surgical conditions.

For any of the methods of application described above, the therapeutic nucleic acid construct is preferably applied to the site of the cancer event (for example, by injection). However, it may also be applied to tissue in the vicinity of the cancer event or to a blood vessel supplying the cells predicted to undergo cancer.

P. Transgenic Animals

Transgenic animals containing mutant, knock-out or modified genes corresponding to the cDNA sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15, or the open reading frame encoding the polypeptide sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16, or fragments thereof having a consecutive sequence of at least about 3, 4, 5, 6, 10, 15, 20, 25, 30, 35 or more amino acid residues, are also included in the invention. Transgenic animals are genetically modified animals into which recombinant, exogenous or cloned genetic material has been experimentally transferred. Such genetic material is often referred to as a "transgene." The nucleic acid sequence of the transgene, in this case a form of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15, may be integrated either at a locus of a genome where that particular nucleic acid sequence is not otherwise normally found or at the normal locus for the transgene. The transgene may consist of nucleic acid sequences derived from the genome of the same species or of a different species than the species of the target animal.

In some embodiments, transgenic animals in which all or a portion of a gene comprising SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 is deleted may be constructed. In those cases where the gene corresponding to SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 or 15 contains one or more introns, the entire gene- all exons, introns and the regulatory sequences- may be deleted. Alternatively, less than the entire gene may be deleted. For example, a single exon and/or intron may be deleted, so as to create an animal expressing a modified version of a protein of the invention.

The term "germ cell line transgenic animal" refers to a transgenic animal in which the genetic alteration or genetic information was introduced into a germ line cell, thereby conferring the ability of the transgenic animal to transfer the genetic information to offspring. If such offspring in fact possess some or all of that alteration or genetic information, then they too are transgenic animals.

The alteration or genetic information may be foreign to the species of animal to which the recipient belongs, foreign only to the particular individual recipient, or may be genetic information already possessed by the recipient. In the last case, the altered or introduced gene may be expressed differently than the native gene.

Transgenic animals can be produced by a variety of different methods including transfection, electroporation, microinjection, gene targeting in embryonic stem cells and recombinant viral and retroviral infection (*see, e.g.*, U.S. Patent No. 4,736,866; U.S. Patent No. 5,602,307; Mullins *et al.* (1993), *Hypertension* 22: 630-633; Brenin *et al.* (1997), *Surg. Oncol.* 6: 99-110; Recombinant Gene Expression Protocols (Methods in Molecular Biology, Vol. 62), Tuan, ed., Humana Press, Totowa, NJ, 1997).

A number of recombinant or transgenic mice have been produced, including those which express an activated oncogene sequence (U.S. Patent No. 4,736,866); express simian SV40 T-antigen (U.S. Patent No. 5,728,915); lack the expression of interferon regulatory factor 1 (IRF-1) (U.S. Patent No. 5,731,490); exhibit dopaminergic dysfunction (U.S. Patent No. 5,723,719); express at least one human gene which participates in blood pressure control (U.S. Patent No. 5,731,489); display greater similarity to the conditions existing in naturally occurring Alzheimer's disease (U.S. Patent No. 5,720,936); have a reduced capacity to mediate cellular adhesion (U.S. Patent No. 5,602,307); possess a bovine growth hormone gene (Clutter *et al.* (1996), *Genetics* 143: 1753-1760); or, are capable of generating a fully human antibody response (McCarthy (1997), *Lancet* 349: 405).

While mice and rats remain the animals of choice for most transgenic experimentation, in some instances it is preferable or even necessary to use alternative animal species. Transgenic procedures have been successfully utilized in a variety of non-murine animals, including sheep, goats, pigs, dogs, cats, monkeys, chimpanzees, hamsters, rabbits, cows and guinea pigs (see, *e.g.*, Kim *et al.* (1997), *Mol. Reprod. Dev.* 46: 515-526; Houdebine (1995), *Reprod. Nutr. Dev.* 35: 609-617; Petters (1994), *Reprod. Fertil. Dev.* 6: 643-645; Schnieke *et al.* (1997), *Science* 278: 2130-2133; and Amoah (1997), *J. Animal Sci.* 75: 578-585).

The method of introduction of nucleic acid fragments into recombination competent mammalian cells can be by any method which favors co-transformation of multiple nucleic acid molecules. Detailed procedures for producing transgenic animals are readily available to one skilled in the art, including the disclosures in U.S. Patent No. 5,489,743 and U.S. Patent No. 5,602,307.

Q. Diagnostic Methods

As the genes and proteins of the invention are differentially expressed in cancerous tissues compared to non-cancerous tissues, the genes and proteins of the invention may be used to diagnose or monitor cancer, to track disease progression, or to differentiate cancerous tissue from non-cancerous tissue samples. One means of diagnosing cancer using the nucleic acid molecules or proteins of the invention involves obtaining tissue from living subjects.

Assays to detect nucleic acid or protein molecules of the invention may be in any available format. Typical assays for nucleic acid molecules include hybridization or PCR based formats. Typical assays for the detection of proteins, polypeptides or peptides of the invention include the use of antibody probes in any available format such as *in situ* binding assays, etc. (see Harlow & Lane, Antibodies - A Laboratory Manual, Cold Spring

Harbor Laboratory Press, Cold Spring Harbor, NY, 1988). In preferred embodiments, assays are carried-out with appropriate controls.

Generally, the diagnostics of the invention can be classified according to whether the embodiment is a nucleic acid or protein-based assay. Some diagnostic assays detect mutations or polymorphisms in the invention nucleic acids or proteins, which contribute to cancerous aberrations. Other diagnostic assays identify and distinguish defects in protein activity by detecting a level of invention RNA or protein in a tested organism that resembles the level of invention RNA or protein in a organism suffering from a disease, such as cancer, or by detecting a level of RNA or protein in a tested organism that is different than an organism not suffering from a disease.

Additionally, the manufacture of kits that incorporate the reagents and methods described in the following embodiments so as to allow for the rapid detection and identification of aberrations in protein activity or level are contemplated. The diagnostic kits can include a nucleic acid probe or an antibody or combinations thereof, which specifically detect a mutant form of the invention protein or a nucleic acid probe or an antibody or combinations thereof, which can be used to determine the level of RNA or protein expression of one or more invention protein. The detection component of these kits will typically be supplied in combination with one or more of the following reagents. A support capable of absorbing or otherwise binding DNA, RNA, or protein will often be supplied. Available supports include membranes of nitrocellulose, nylon or derivatized nylon that can be characterized by bearing an array of positively charged substituents. One or more restriction enzymes, control reagents, buffers, amplification enzymes, and non-human polynucleotides like calf-thymus or salmon-sperm DNA can be supplied in these kits.

Useful nucleic acid-based diagnostic techniques include, but are not limited to, direct DNA sequencing, gradient gel electrophoresis, Southern Blot analysis, single-stranded confirmation analysis (SSCA), RNAse protection assay, dot blot analysis, nucleic

acid amplification, allele-specific PCR and combinations of these approaches. The starting point for these analyses is isolated or purified nucleic acid from a biological sample. It is contemplated that tissue biopsies would provide a good sample source. The nucleic acid is extracted from the sample and can be amplified by a DNA amplification technique such as the Polymerase Chain Reaction (PCR) using primers. Those of skill in the art will readily recognize methods available for confirming the presence of polymorphisms. In addition, any addressable array technology known in the art can be employed with this aspect of the invention. One particular embodiment of polynucleotide arrays is known as Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092.

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid assays. There are several ways to produce labeled nucleic acids for hybridization or PCR including, but not limited to, oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, a nucleic acid encoding an invention protein can be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and can be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides. A number of companies such as Pharmacia Biotech (Piscataway, NJ), Promega (Madison, WI), and U.S. Biochemical Corp (Cleveland, OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as, substrates, cofactors, inhibitors, magnetic particles and the like.

In preferred protein-based diagnostic, antibodies of the invention are attached to a support in an ordered array wherein a plurality of antibodies are attached to distinct regions of the support that do not overlap with each other. Those of skill in the art will readily recognize available assays that are protein-based diagnostics. Proteins are obtained from biological samples and are labeled by conventional approaches (e.g.,

radioactivity, colorimetrically, or fluorescently). Employing labeled standards of a known concentration of mutant and/or wild-type invention protein, an investigator can accurately determine the concentration of the invention protein in a sample and from this information can assess the expression level of the particular form of the protein. Conventional methods in densitometry can also be used to more accurately determine the concentration or expression level of such protein. These approaches are also easily automated using technology known to those of skill in the art of high throughput diagnostic analysis. As detailed above, any addressable array technology known in the art can be employed with this aspect of the invention and display the protein arrays on the chips in an attempt to maximize antibody binding patterns and diagnostic information.

As discussed above, the presence or detection of a polymorphism in an invention gene or protein can provide a diagnosis of a cancer or similar malady in an organism. Additional embodiments include the preparation of diagnostic kits comprising detection components, such as antibodies, specific for a particular polymorphic variant of invention gene or protein. The detection component will typically be supplied in combination with one or more of the following reagents. A support capable of absorbing or otherwise binding RNA or protein will often be supplied. Available supports for this purpose include, but are not limited to, membranes of nitrocellulose, nylon or derivatized nylon that can be characterized by bearing an array of positively charged substituents, and Genechips™ or their equivalents. One or more enzymes, such as Reverse Transcriptase and/or Taq polymerase, can be furnished in the kit, as can dNTPs, buffers, or non-human polynucleotides like calf-thymus or salmon-sperm DNA. Results from the kit assays can be interpreted by a healthcare provider or a diagnostic laboratory. Alternatively, diagnostic kits are manufactured and sold to private individuals for self-diagnosis.

In addition to diagnosing disease according to the presence or absence of a polymorphism, some diseases involving cancer result from skewed levels of invention protein or gene in particular tissues or aberrant patterns of invention protein expression. By monitoring the level of expression in various tissues, for example, a diagnosis can be

made or a disease state can be identified. Similarly, by determining ratios of the level of expression of various invention proteins in specific tissues (e.g., patterns of expression) a prognosis of health or disease can be made. The levels of invention protein expression in various tissues from healthy individuals, as well as, individuals suffering from cancers is determined. These values can be recorded in a database and can be compared to values obtained from tested individuals. Additionally, the ratios or patterns of expression in various tissues from both healthy and diseased individuals is recorded in a database. These analyses are referred to as "disease state profiles" and by comparing one disease state profile (e.g. from a healthy or diseased individual) to a disease state profile from a tested individual, a clinician can rapidly diagnose the presence or absence of disease.

The nucleic acid and protein-based diagnostic techniques described above can be used to detect the level or amount or ratio of expression of invention genes or proteins in a tissue. Through quantitative Northern hybridizations, *in situ* analysis, immunohistochemistry, ELISA, genechip array technology, PCR, and Western blots, for example, the amount or level of expression of RNA or protein for a particular invention protein (wild-type or mutant) can be rapidly determined and from this information ratios of expression can be ascertained. Alternatively, the invention proteins to be analyzed can be family members that are currently unknown but which are identified based on their possession of one or more of the homology regions described above.

Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the compounds of the present invention and practice the claimed methods. The following working examples therefore, specifically point out preferred embodiments of the present invention, and are not to be construed as limiting in any way the remainder of the disclosure.

EXAMPLES

EXAMPLE 1: Identification of Differentially Expressed mRNA in Cancers - 1

Global changes in gene expression between tumor biopsies and normal tissues have been examined using the GeneExpress Oncology Datasuite™ of Gene Logic, Inc. (Gaithersburg, MD). The database includes the gene expression profiles, generated by using the Affymetrix Human Genome U95 array, derived from normal and cancer tissue samples from many different organs. Among the tissue samples in the database, applicants analyzed the expression profiles of normal and cancer tissue sets from breast, colon, esophagus, kidney, liver, lung, lymph node, ovary, pancreas, prostate, rectum, and stomach.

The Affymetrix Human Genome U95 array contains 63,175 probe sets. A probe set is a set of probes to detect one transcript (a gene or a cDNA clone), and usually consists of 16-20 oligonucleotide probe pairs. These probe pairs include perfectly matched sets and mismatched sets, both of which are necessary for the calculation of average difference. Average difference serves as a relative indicator of the level of expression of a transcript and is a measure of the intensity difference for each probe pair, calculated by subtracting the intensity of the mismatch from the intensity of the perfect match. This takes into consideration variability in hybridization among probe pairs and other hybridization artifacts that could affect the fluorescence intensities. Using the average difference value that has been calculated, an absolute call for each gene is made; "Absent" (= not detected), "Present" (= detected) or "Marginal" (= not clearly Absent or Present).

Differential expression of genes between cancerous and normal tissue samples was determined with the following statistical methods. (1) For each probe set, average difference values and absolute calls were determined by Affymetrix Microarray Suite (v4.0). (2) In a given sample set, outliers among the tissue samples were detected by Principal Component Analysis (PCA) using MatLab program (The MathWorks, Inc., Natick, MA). The data points used in the PCA were the average differences of randomly selected probe sets (5,000~6,000 probe sets). Outliers were excluded from further analysis.

(3) Variations of gene expression were analyzed by using the Fold Change Analysis tool of GeneExpress program. The fold change (cancerous/normal) was calculated by comparing the mean average difference for each gene in a cancerous sample set against the mean average difference of that gene in the normal tissue sample set. Genes showing at least 3-fold increases or decreases in expression level were obtained. Genes were included in the analysis if they had a p-value of less than or equal to 0.05 as determined by an Analysis of Variance Test (Steel *et al.*, Principles and Procedures of Statistics: A Biometrical Approach, Third Ed., McGraw-Hill, 1997). (4) Genes showing differential expression in at least 5 different cancer types were selected.

Analysis of the chip data showed that the expression of the marker LFG1 was significantly up-regulated in cancer tissue samples compared to samples from normal tissue. The expression level of LFG1 (SEQ ID NO: 1 or 3) can be measured by chip sequence fragment no. 91875_s_at on Affymetrix GeneChips® U95. The 91875_s_at sequence is derived from the EST AI053741. The expression levels of 91875_s _at in various malignant neoplasms, compared to normal control tissues, are shown in Table 1, where the fold-change, the direction of the change (up- or down-regulation), p-value are also indicated. The fold change (cancerous/normal) was calculated by comparing the geometric mean of average difference in a cancerous sample set against the geometric mean of average difference in the normal tissue sample set. A fold change greater than 1.5 was considered to be significant (Wodicka *et al.* (1997), *Nature Biotech.* 15:1359-1367). Also indicated in the Table 1 are, for each tissue type, the numbers of samples that are called present, absent, or marginal together with the total number of samples in that sample set. These data indicate that up-regulation of LFG1 may be diagnostic for cancer.

TABLE 1

Tissue	Pathology / Morphology	Geometric Mean	Number of Samples				Fold Change	Direction	p-value
			Total	Present	Marginal	Absent			
BREAST	NORMAL TISSUE, NOS	22.71	34	8	4	22			
	INFILTRATING DUCT CARCINOMA	184.04	61	61	0	0	8.11	up	0
	INFILTRATING LOBULAR CARCINOMA	104.36	10	9	0	1	4.60	up	0.00456
COLON	NORMAL TISSUE, NOS	76.46	24	23	0	1			
	ADENOCARCINOMA, NOS	244.76	36	35	0	1	3.20	up	0.00001
ESOPHAGUS	NORMAL TISSUE, NOS	50.47	18	16	1	1			
	ADENOCARCINOMA, NOS	297.56	8	8	0	0	5.90	up	0.00367
KIDNEY	NORMAL TISSUE, NOS	20.00	25	1	0	24			
	CLEAR CELL CARCINOMA	60.48	11	10	1	0	3.02	up	0.00082
	RENAL CELL CARCINMA	65.01	16	13	0	3	3.25	up	0.00011
LIVER	NORMAL TISSUE, NOS	22.06	19	3	0	16			
	HEPATOCELLULAR CARCINOMA, NOS	86.74	23	21	0	2	3.93	up	0
LUNG	NORMAL TISSUE, NOS	21.27	32	6	0	26			
	ADENOCARCINOMA, NOS	122.81	39	38	0	1	5.77	up	0
OVARY	NORMAL TISSUE, NOS	20.21	23	0	0	23			
	PAPILLARYSEROUSADENOCARCINOMA	112.80	23	21	0	2	5.58	up	0
PANCREAS	NORMAL TISSUE, NOS	20.02	20	1	0	19			
	ADENOCARCINOMA, NOS	72.55	25	22	0	3	3.62	up	0
RECTUM	NORMAL TISSUE, NOS	78.86	20	20	0	0			
	ADENOCARCINOMA, NOS	259.95	22	22	0	0	3.30	up	0.00008
STOMACH	NORMAL TISSUE, NOS	36.06	18	7	0	11			
	ADENOCARCINOMA, NOS	218.74	38	36	0	2	6.07	up	0

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 91875_s _at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer). PCR primers (5'-GCTGAAGCAGGAAAATCGCTT-3' (SEQ ID NO: 17) and 5'-TGAGACGGAGTCTCACTCGGT-3' (SEQ ID NO: 18)) designed based on the sequence information file of the specific Affymetrix fragment (91875_s _at) were used in the assay. The target gene in each RNA sample (10 ng of total RNA) was assayed relative to a reference gene. For this purpose, primers (5'-GTTTTTCCTAATTTTGGCATGAAC-3' (SEQ ID NO: 19) and 5'-CGCCCAAGCTTTTCCTTTT-3' (SEQ ID NO: 20)) specific to the CTBP1 gene (C-terminal binding protein 1) were used to serve as control primers. This approach provides the relative expression as measured by cycle threshold (Ct) value of the target mRNA relative to an amount of CTBP1 Ct value. The sample panel included total RNA pairs of normal and tumor tissues from colon, kidney, liver, lung, ovary, stomach and pancreas (Ambion, Inc., Austin, TX). The Q-RT-PCR data confirms the up-regulation of LFG1 in cancer compared to normal samples.

EXAMPLE 2: Cloning of Full-Length Human cDNA (LFG1) Corresponding to Differentially Expressed mRNA Species

The full-length cDNA having SEQ ID NO: 1 or 3 was obtained by polymerase chain reaction (PCR) and rapid amplification of cDNA ends (RACE) using cDNA library from human heart (ResGen, Huntsville, AL). Gene-specific oligos for PCR (5'-CACCCTTTGCCTCTGTCACTTCCGCA-3' (SEQ ID NO: 21), 5'-GCTGGAGCACCAGGACTGCATTG-3' (SEQ ID NO: 22), 5'-GGAGCTGAGCAGCAGTGTAATGAA-3' (SEQ ID NO: 23), 5'-GAGGCCTGCCTGAAGGAGGAGCTTC-3' (SEQ ID NO: 24), 5'-TCTGGAAGTAGTGCAGACGCCTCAGG-3' (SEQ ID NO: 25), 5'-AGCCAACGTCGGCTTTGTTATCCAGC-3' (SEQ ID NO: 26), 5'-

GCTGTCAGATATGATGGTTCTGGAC-3' (SEQ ID NO: 27), 5'-CCAGCCTCACCACTGTTGGGTTGC-3' (SEQ ID NO: 28), 5'-CATTCTCTGAGCTGTATTAGTGT-3' (SEQ ID NO: 29), 5'-CCTGAGCTGGAATGACCTGCA-3' (SEQ ID NO: 30), 5'-CTTTGTGTTGGCTGCAGCCACA-3' (SEQ ID NO: 31), 5'-TGAGGAGAGACTTTGCTGACTGGT-3' (SEQ ID NO: 32), 5'-GTCCTGTCTGGCGGTGCCGA-3' (SEQ ID NO: 33), 5'-GCTCCAGGATCCCCTGTCACCTGGGCCTTCTGCCTTTTGGCT-3' (SEQ ID NO: 34), 5'-CCATATGGAGAGGAGAGCAGCGGGCCCA-3' (SEQ ID NO: 35), 5'-GAAGGAGGAACATGGAGAGGAGA-3' (SEQ ID NO: 36), 5'-CCATATGCCCCGGGTAGTCTACTGCAT-3' (SEQ ID NO: 37), and 5'-GTCGACTCGAGTCACTTCCGCAAAAACTTCTTG-3' (SEQ ID NO: 38)) and RACE (5'-TCCATTCCGAAGGCTCTCCTCC-3' (SEQ ID NO: 39), 5'-GTCTGTGTGACGGAAATGTAAGC-3' (SEQ ID NO: 40), and 5'-GAAGGTCGAAGGCAGACCGATGT-3' (SEQ ID NO: 41)) were designed based on predicted genes containing the 91875_s_at sequence using Human Genome Browser (University of California, Santa Cruz). The amplified products with the primers were incorporated into PCR4-Topo vector using Topo Cloning System (Invitrogen, Carlsbad, CA), and followed by sequencing.

The nucleotide sequence of the full-length human cDNAs corresponding to the differentially regulated mRNA detected above is set forth in SEQ ID NOS: 1 and 3. In the former, the cDNA comprises 5293 base pairs. In the latter, the cDNA comprises 5317 base pairs.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 1, at nucleotides 390-4880 (390-4883 including the stop codon), encodes a protein of 1497 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 1 is set forth in SEQ ID NO: 2. Figure 2 shows the results of a hydrophobicity

analysis of the amino acid sequence of SEQ ID NO: 2 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.* 157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 3, at nucleotides 12-4904 (12-4907 including the stop codon), encodes a protein of 1631 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 3 is set forth in SEQ ID NO: 4. Figure 3 shows the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NO: 4 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.* 157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above.

The protein sequence of SEQ ID NO: 2 is identical to that of SEQ ID NO: 4, except that SEQ ID NO: 2 lacks the first 134 amino acids at the N-terminus of SEQ ID NO: 4.

SEQ ID NOS: 2 and 4 contain Calponin homology domain (amino acid positions 38-145 of SEQ ID NO: 4), IQ domain for calmodulin-binding (amino acid positions 629-646 of SEQ ID NO: 2 and amino acid positions 763-780 of SEQ ID NO: 4), RasGAP domain (amino acid positions 858-1195 of SEQ ID NO: 2 and amino acid positions 992-1329 of SEQ ID NO: 4), and RasGAP C-terminal domain (amino acid positions 1298-1421 of SEQ ID NO: 2 and amino acid positions 1432-1555 of SEQ ID NO: 4). SEQ ID NOS: 2 and 4 are similar to IQGAP proteins (Weissbach *et al.* (1994), *J Biol Chem* 269:20517-20521; Brill *et al.* (1996), *Mol Cell Biol* 16:4869-4878). IQGAP binds to and modulate the function of proteins involved in cytoskeletal structure, cell-cell adhesion, and proliferation signaling (Fukada *et al.* (2002), *Cell* 109: 1-20; Briggs *et al.* (2002), *J Biol Chem* 277: 7453-7465; McCallum *et al.* (1998), *J Biol Chem* 273: 22537-22544). IQGAP1-deficient mice exhibited a significant increase in late-onset gastric hyperplasia relative to wild-type (Li *et al.* (2000), *Mol Cell Biol* 20: 697-701).

Analysis by Northern blot was performed to determine the size of the mRNA transcripts that correspond to LFG1. A Northern blot containing total RNAs from various human tissues was used (Human 12-Lane MTN Blot, Clontech, Palo Alto, CA), and an EST containing 91875_s _at sequence was radioactively labeled by the random primer method and used to probe the blot. The blot was hybridized in 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhart's solution, and 0.2 mg/ml herring sperm DNA at 42°C and washed with 0.2X SSC containing 0.1% SDS at room temperature. The Northern blot showed three transcripts for this gene, which are approximately 7.2 kb, and 6.3 kb in size. This corresponds to the sizes of the LFG1 clones (SEQ ID NO: 1 and 3).

EXAMPLE 3: Identification of Differentially Expressed mRNA in Cancers - 2

The process in EXAMPLE 1 was repeated except that the marker LFG2 was used instead of the marker LFG1.

Analysis of the chip data showed that the expression of the marker LFG2 was significantly down-regulated in cancer tissue samples compared to samples from normal tissue. The expression level of LFG2 (SEQ ID NO: 5) can be measured by chip sequence fragment no. 82941_at on Affymetrix GeneChips® U95. The 82941_at sequence is derived from the EST AI277612. The expression levels of 82941_at in various malignant neoplasms, compared to normal control tissues, are shown in Table 2, where the fold-change, the direction of the change (up- or down-regulation), p-value are also indicated. The fold change (cancerous/normal) was calculated by comparing the geometric mean of average difference in a cancerous sample set against the geometric mean of average difference in the normal tissue sample set. A fold-change greater than 1.5 was considered to be significant (Wodicka et al. (1997), Nature Biotech. 15:1359-1367). Also indicated in the Table 2 are, for each tissue type, the numbers of samples that are called present, absent, or marginal together with the total number of samples in that sample set. These data indicate that down-regulation of LFG2 may be diagnostic for cancer.

TABLE 2

Tissue	Pathology / Morphology	Geometric Mean	Number of Samples				Fold Change	Direction	p-value
			Total	Present	Marginal	Absent			
BREAST	NORMAL TISSUE, NOS	1147.66	34	34	0	0			
	INFILTRATING DUCT CARCINOMA	129.77	61	26	3	32	8.71	down	0
	INFILTRATING LOBULAR CARCINOMA	183.37	10	6	1	3	5.48	down	0.00002
COLON	NORMAL TISSUE, NOS	890.06	24	23	1	0			
	ADENOCARCINOMA, NOS	163.35	36	17	1	18	5.39	down	0
ESOPHAGUS	NORMAL TISSUE, NOS	612.34	18	18	0	0			
	ADENOCARCINOMA, NOS	265.11	8	7	1	0	2.31	down	0.02218
LIVER	NORMAL TISSUE, NOS	182.73	19	11	1	7			
	HEPATOCELLULAR CARCINOMA, NOS	114.69	23	7	1	15	1.51	down	0.01211
LUNG	NORMAL TISSUE, NOS	535.64	32	30	2	0			
	ADENOCARCINOMA, NOS	119.36	39	17	3	19	4.27	down	0
LYMPH NODE	NORMAL TISSUE, NOS	454.08	9	7	0	2			
	MALIGNANT LYMPHOMA, NOS	123.13	12	5	0	7	3.24	down	0.02245
OVARY	NORMAL TISSUE, NOS	279.99	23	21	0	2			
	PAPILLARY SEROUS ADENOCARCINOMA	85.45	23	7	1	15	3.5	down	0
PROSTATE	NORMAL TISSUE, NOS	195.77	19	13	1	5			
	ADENOCARCINOMA, NOS	80.06	19	2	2	15	2.57	down	0.00011
RECTUM	NORMAL TISSUE, NOS	943.86	20	19	0	1			
	ADENOCARCINOMA, NOS	176.45	22	13	2	7	5.2	down	0
STOMACH	NORMAL TISSUE, NOS	414.40	18	16	0	2			
	ADENOCARCINOMA, NOS	125.39	38	17	2	19	3.21	down	0

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 82941_at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer). PCR primers (5'-GAATGTGTCAGAGACAAGTGCAGC-3' (SEQ ID NO: 42) and 5'-TGTAGAAACTCTTGGACTAATGGAGG-3' (SEQ ID NO: 43)) designed based on the sequence information file of the EST containing the Affymetrix fragment (82941_at) were used in the assay. The target gene in each RNA sample (10 ng of total RNA) was assayed relative to a reference gene. For this purpose, primers (5'-GTTTTTCCTAATTTTGGCATGAAC-3' (SEQ ID NO: 19) and 5'-CGCCCAAGCTTTTCCTTTT-3' (SEQ ID NO: 20)) specific to the CTBP1 gene (C-terminal binding protein 1) were used to serve as control primers. This approach provides the relative expression as measured by cycle threshold (Ct) value of the target mRNA relative to an amount of CTBP1 Ct value. The sample panel included total RNA pairs of normal and tumor tissues from colon, liver, lung, ovary, and stomach (Ambion, Inc., Austin, TX). The Q-RT-PCR data confirms the down-regulation of LFG2 in cancer compared to normal samples.

EXAMPLE 4: Cloning of Full-Length Human cDNA (LFG2) Corresponding to Differentially Expressed mRNA Species

The full-length cDNA having SEQ ID NO: 5 was obtained by the oligo-pulling method using the GeneTrapper assay (Life Technologies, Rockville, MD). Briefly, a gene-specific oligo (5'-GAATGTGTCAGAGACAAGTGCAGC-3' (SEQ ID NO: 42)) was designed based on the sequence of the EST containing 82941_at sequence. The oligo was labeled with biotin and used to hybridize with 5 μg of single strand plasmid DNA (cDNA recombinants) from a poorly differentiated stomach adenocarcinoma library (NCI CGAP Gas4) (ResGen, Huntsville, AL) following the procedures of Sambrook *et al.* The hybridized cDNAs were separated by streptavidin-conjugated beads and eluted by heating.

The eluted cDNA was converted to double strand plasmid DNA and used to transform *E. coli* cells (DH10B) and the longest cDNA was screened. After positive selection was confirmed by PCR using gene-specific primers, the cDNA clone was subjected to DNA sequencing.

The nucleotide sequence of the full-length human cDNAs corresponding to the differentially regulated mRNA detected above is set forth in SEQ ID NO: 5. The cDNA comprises 3608 base pairs.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 5, at nucleotides 424-1908 (424-1911 including the stop codon), encodes a protein of 495 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 5 is set forth in SEQ ID NO: 6.

SEQ ID NO: 6 has homology to scavenger receptors, which are involved in endocytosis of selected polyanionic ligands, phagocytosis of apoptotic cells and bacteria, cell adhesion, and development of atherosclerosis (Peiser *et al.* (2002), *Curr. Opin. Immunol.* 14:123-128; Resnick *et al.* (1994), *Trends Biol. Sci.* 19:5-8). Based on published studies of scavenger receptors, SEQ ID NO: 6 contains a cytoplasmic domain (amino acid positions 1-35), a transmenbrane domain (amino acid positions 36-58), an α-helical coiled-coil domain (amino acid positions 90-301), a collagen-like domain (amino acid positions 305-380), and a scavenger receptor cystein-rich (SRCR) domain (amino acid positions 393-493). The SRCR domain contains six cysteine residues (amino acid positions 418, 431, 462, 472, 482, and 492), which may participate in intradomain disulfide bonds. SEQ ID NO: 6 also exhibits homology to a mouse homologue (GenBank Accession No. BC016096). It shows 70% identity over the entire contiguous sequence.

Figure 4 shows the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NO: 6 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.*

157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above.

Analysis by Northern blot was performed to determine the size of the mRNA transcripts that correspond to LFG2. A Northern blot containing total RNAs from various human tissues was used (Human MTN Blot, Clontech, Palo Alto, CA), and the EST containing 82941_at sequence was radioactively labeled by the random primer method and used to probe the blot. The blot was hybridized in 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhart's solution, and 0.2 mg/ml herring sperm DNA at 42°C and washed with 0.2X SSC containing 0.1% SDS at room temperature. The Northern blot showed a single transcript for this gene, which is approximately 3.7 kb in size. This corresponds to the size of the LFG2 clone (SEQ ID NO: 5).

EXAMPLE 5: Identification of Differentially Expressed mRNA in Cancers - 3

The process in EXAMPLE 1 was repeated except that the marker LFG3 was used instead of the marker LFG1.

Analysis of the chip data showed that the expression of the marker LFG3 was significantly down-regulated in cancer tissue samples compared to samples from normal tissue. The expression level of LFG3 (SEQ ID NO: 7) can be measured by chip sequence fragment no. 46104_at on Affymetrix GeneChips® U95. The 46104_at sequence is derived from the EST AA772055. The expression levels of 46104_at in various malignant neoplasms, compared to normal control tissues, are shown in Table 3, where the fold-change, the direction of the change (up- or down-regulation), p-value are also indicated. The fold change (cancerous/normal) was calculated by comparing the geometric mean of average difference in a cancerous sample set against the geometric mean of average difference in the normal tissue sample set. A fold-change greater than 1.5 was considered to be significant (Wodicka et al. (1997), Nature Biotech. 15:1359-1367). Also indicated in the

Table 3 are, for each tissue type, the numbers of samples that are called present, absent, or marginal together with the total number of samples in that sample set. These data indicate that down-regulation of LFG3 may be diagnostic for cancer.

TABLE 3

Tissue	Pathology / Morphology	Geometric Mean	Number of Samples				Fold Change	Direction	p-value
			Total	Present	Marginal	Absent			
BREAST	NORMAL TISSUE, NOS	64.52	34	31	0	3			
	INFILTRATING DUCT CARCINOMA	27.24	61	18	1	42	2.25	down	0
	INFILTRATING LOBULAR CARCINOMA	29.52	10	4	0	6	2.21	down	0.00004
COLON	NORMAL TISSUE, NOS	315.46	24	24	0	0			
	ADENOCARCINOMA, NOS	102.99	36	31	0	5	3.02	down	0.00016
ESOPHAGUS	NORMAL TISSUE, NOS	272.48	18	17	0	1			
	ADENOCARCINOMA, NOS	41.25	8	6	0	2	6.60	down	0.00001
KIDNEY	NORMAL TISSUE, NOS	2626.88	25	25	0	0			
	CLEAR CELL ADENOCARCINOMA, NOS	344.66	11	11	0	0	7.62	down	0.00003
	RENAL CELL CARCINOMA	355.71	16	14	0	2	7.38	down	0.00005
OVARY	NORMAL TISSUE, NOS	1098.41	23	23	0	0			
	PAPILLARY SEROUS ADENOCARCINOMA	178.15	23	22	0	1	6.17	down	0
PROSTATE	NORMAL TISSUE, NOS	274.49	19	19	0	0			
	ADENOCARCINOMA, NOS	117.26	19	18	0	1	2.34	down	0.00016
RECTUM	NORMAL TISSUE, NOS	410.22	20	20	0	0			
	ADENOCARCINOMA, NOS	72.98	22	16	0	6	5.38	down	0
STOMACH	NORMAL TISSUE, NOS	71.10	18	10	0	8			
	ADENOCARCINOMA, NOS	35.49	38	15	1	22	1.96	down	0.00459

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 46104_at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer). PCR primers (5'-GTATGCATCAGAATTCCCTATAGATCTTT-3' (SEQ ID NO: 44) and 5'-TAGATGTTTGGGCAACAGCCT-3' (SEQ ID NO: 45)) designed based on the sequence information file of the EST containing the Affymetrix fragment (46104_at) were used in the assay. The target gene in each RNA sample (10 ng of total RNA) was assayed relative to a reference gene. For this purpose, primers (5'-GTTTTTCCTAATTTTGGCATGAAC-3' (SEQ ID NO: 19) and 5'-CGCCCAAGCTTTTCCTTTT-3' (SEQ ID NO: 20)) specific to the CTBP1 gene (C-terminal binding protein 1) were used to serve as control primers. This approach provides the relative expression as measured by cycle threshold (Ct) value of the target mRNA relative to an amount of CTBP1 Ct value. The sample panel included total RNA pairs of normal and tumor tissues from colon, kidney, ovary, pancreas, and stomach (Ambion, Inc., Austin, TX). The Q-RT-PCR data confirms the down-regulation of LFG3 in cancer compared to normal samples.

EXAMPLE 6: Cloning of Full-Length Human cDNA (LFG3) Corresponding to Differentially Expressed mRNA Species

The full-length cDNA having SEQ ID NO: 7 was obtained by the oligo-pulling method using the GeneTrapper assay (Life Technologies, Rockville, MD). Briefly, a gene-specific oligo (5'-GTATGCATCAGAATTCCCTATAGATCTTT-3' (SEQ ID NO: 44)) was designed based on the sequence of the EST containing 46104_at sequence. The oligo was labeled with biotin and used to hybridize with 5 μg of single strand plasmid DNA (cDNA recombinants) from human fetal kidney (ResGen, Huntsville, AL) following the procedures of Sambrook *et al.* The hybridized cDNAs were separated by streptavidin-conjugated beads and eluted by heating. The eluted cDNA was converted to double strand plasmid DNA and used to transform *E. coli* cells (DH10B) and the longest cDNA was

screened. After positive selection was confirmed by PCR using gene-specific primers, the cDNA clone was subjected to DNA sequencing. The 5'-end of LFG3 was identified by rapid amplification of cDNA ends (RACE) using the cDNA prepared from human fetal kidney (Clontech, Palo Alto, CA) and a gene specific primer (5'-TTCCTTCACCAAAGGCATCCAGCCATTCTATG-3' (SEQ ID NO: 46)).

The nucleotide sequence of the full-length human cDNAs corresponding to the differentially regulated mRNA detected above is set forth in SEQ ID NO: 7. The cDNA comprises 3162 base pairs.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 7, at nucleotides 405-1835 (405-1838 including the stop codon), encodes a protein of 477 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 7 is set forth in SEQ ID NO: 8.

SEQ ID NO: 8 is similar to monocarboxylate transporters (MCTs) and contains ten predicted transmembrane domains (amino acids positions 10-29, 80-99, 107-128, 140-160, 274-295, 312-332, 339-360, 363-384, 396-416, and 433-451). MCT proteins catalyze the facilitated transport of monocarboxylates such as lactate, pyruvate, branched-chain oxo acids, ketone bodies, beta-hydroxy-butylate, and acetate (Halestrap and Price (1999), *Biochem. J.* 343:281-299). Table 4 summarizes the similarity ratios of SEQ ID NO: 4 with the eight known monocarboxylate transporters.

TABLE 4. Homology of LFG3 with MCT proteins

Protein	Size (amino acids)	Identity (%)	Positives (%)
MCT1	500	17.5	34.3
MCT2	478	19.5	35.5
MCT3	504	19.5	34.1
MCT4	465	19.0	33.2

MCT5	487	22.1	36.9
MCT6	505	16.4	31.5
MCT7	523	20.1	35.2
MCT8	613	15.9	27.9

Figure 5 shows the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NO: 8 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.* 157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above.

Analysis by Northern blot was performed to determine the size of the mRNA transcripts that correspond to LFG3. A Northern blot containing total RNAs from various human tissues was used (Human 12-Lane MTN Blot, Clontech, Palo Alto, CA), and the EST containing 46104_at sequence was radioactively labeled by the random primer method and used to probe the blot. The blot was hybridized in 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhart's solution, and 0.2 mg/ml herring sperm DNA at 42°C and washed with 0.2X SSC containing 0.1% SDS at room temperature. The Northern blot showed a single transcript for this gene, which is approximately 4.2 kb in size. This corresponds to the size of the LFG3 clone (SEQ ID NO: 7).

EXAMPLE 7: Identification of Differentially Expressed mRNA in Cancers - 4

The process in EXAMPLE 1 was repeated except that the marker LFG4 was used instead of the marker LFG1.

Analysis of the chip data showed that the expression of the marker LFG4 was significantly down-regulated in cancer tissue samples compared to samples from normal tissue. The expression level of LFG4 (SEQ ID NO: 9) can be measured by chip sequence fragment no. 62158_at on Affymetrix GeneChips® U95. The 622158_at sequence is derived from the EST AI123532. The expression levels of 62158_at in various malignant

neoplasms, compared to normal control tissues, are shown in Table 5, where the fold-change, the direction of the change (up- or down-regulation), p-value are also indicated. The fold change (cancerous/normal) was calculated by comparing the geometric mean of average difference in a cancerous sample set against the geometric mean of average difference in the normal tissue sample set. A fold-change greater than 1.5 was considered to be significant (Wodicka et al. (1997), Nature Biotech. 15:1359-1367). Also indicated in the Table 5 are, for each tissue type, the numbers of samples that are called present, absent, or marginal together with the total number of samples in that sample set. These data indicate that down-regulation of LFG4 may be diagnostic for cancer.

TABLE 5

Tissue	Pathology / Morphology	Geometric Mean	Number of Samples				Fold Change	Direction	p-value
			Total	Present	Marginal	Absent			
BREAST	NORMAL TISSUE, NOS	156.75	34	33	0	1			
	INFILTRATING DUCT CARCINOMA	90.09	61	51	0	10	1.74	down	0.00001
COLON	NORMAL TISSUE, NOS	234.06	24	22	2	0			
	ADENOCARCINOMA, NOS	64.02	36	24	0	12	3.66	down	0
KIDNEY	NORMAL TISSUE, NOS	134.17	25	23	0	2			
	CLEAR CELL ADENOCARCINOMA, NOS	78.59	11	7	1	3	1.71	down	0.08272
	RENAL CELL CARCINOMA	55.31	16	9	0	7	2.43	down	0.0021
LUNG	NORMAL TISSUE, NOS	179.71	32	32	0	0			
	ADENOCARCINOMA, NOS	47.39	39	17	3	19	3.79	down	0
LYMPH NODE	NORMAL TISSUE, NOS	140.51	9	7	1	1			
	MALIGNANT LYMPHOMA, NOS	41.43	12	5	1	6	3.39	down	0.00207
OVARY	NORMAL TISSUE, NOS	125.19	23	21	0	2			
	PAPILLARY SEROUS ADENOCARCINOMA	37.23	23	4	0	19	3.36	down	0
PROSTATE	NORMAL TISSUE, NOS	191.94	19	18	0	1			
	ADENOCARCINOMA, NOS	103.47	19	16	0	3	1.86	down	0.00185
RECTUM	NORMAL TISSUE, NOS	317.95	20	20	0	0			
	ADENOCARCINOMA, NOS	74.28	22	16	1	5	4.28	down	0
STOMACH	NORMAL TISSUE, NOS	161.77	18	17	0	1			
	ADENOCARCINOMA, NOS	84.55	38	27	2	9	1.91	down	0.0062

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 62158_at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer). PCR primers (5'-AAATGTCTGATTACCCCATTTTATCAGT-3' (SEQ ID NO: 47) and 5'-TAATCCTGAAATGAACAGCTAACA-3') (SEQ ID NO: 48) designed based on the sequence information file of the EST containing the Affymetrix fragment (62158_at) were used in the assay. The target gene in each RNA sample (10 ng of total RNA) was assayed relative to a reference gene. For this purpose, primers (5'-GTTTTTCCTAATTTTGGCATGAAC-3' (SEQ ID NO: 19) and 5'-CGCCCAAGCTTTTCCTTTT-3' (SEQ ID NO: 20)) specific to the CTBP1 gene (C-terminal binding protein 1) were used to serve as control primers. This approach provides the relative expression as measured by cycle threshold (Ct) value of the target mRNA relative to an amount of CTBP1 Ct value. The sample panel included total RNA pairs of normal and tumor tissues from colon, liver, lung, ovary, pancreas, and stomach (Ambion, Inc., Austin, TX). The Q-RT-PCR data confirms the down-regulation of LFG4 in cancer compared to normal samples.

EXAMPLE 8: Cloning of Full-Length Human cDNA (LFG4) Corresponding to Differentially Expressed mRNA Species

The full-length cDNA having SEQ ID NO: 9 was obtained by rapid amplification of cDNA ends (RACE). Briefly, gene-specific oligos (5'-TAATGTTAGAGTAACAGCATTTTCCTTCAA-3' (SEQ ID NO: 49) and 5'-TGCCCCACACTAACTCAGTTCTTGTGATG-3' (SEQ ID NO: 50)) were designed based on the sequence of the EST containing 62158_at sequence. The oligos was used for PCR amplification of the cDNAs prepared from human brain (Clontech, Palo Alto, CA). The amplified products with the primers were incorporated into PCR4-Topo vector using Topo Cloning System (Invitrogen, Carlsbad, CA), and followed by sequencing.

The nucleotide sequence of the full-length human cDNAs corresponding to the differentially regulated mRNA detected above is set forth in SEQ ID NO: 9. The cDNA comprises 4891 base pairs.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 9, at nucleotides 89-1150 (89-1153 including the stop codon), encodes a protein of 354 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 9 is set forth in SEQ ID NO: 10.

SEQ ID NO: 10 is similar to rat Kilon and chicken Neurotractin (Funatsu *et al.* (1999), *J Biol Chem* 274:8224-8230; Marg *et al.* (1999), *J Cell Biol* 145:865-876). Protein sequence analysis reveals a secretory signal peptide (amino acid positions 1-33), three immunoglobulin domains (amino acid positions 47-136, 145-208, and 231-312), and six putative *N*-linked glycosylation sites (amino acid positions 73, 155, 275, 286, 294, and 307). Kilon/Neurotractin is a member of IgLON subfamily of the immunoglobulin superfamily. IgLONs are a family of glycosylphosphatidylinositol (GPI)-linked cell adhesion molecules which are thought to modify neurite outgrowth and might play a role in cell-cell adhesion and recognition (Miyate *et al.* (2000), *J Comparative Neurol* 424:74-85).

Figure 6 shows the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NO: 10 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.* 157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above. This hydropathy plot shows the presence of hydrophobic region at the C-terminus. In case of GPI-anchored proteins, the addition of the GPI anchor is known to occur after the cleavage of the C-terminal hydrophobic region. A putative GPI anchor attachment site was found (Gly at the amino acid position 324).

Analysis by Northern blot was performed to determine the size of the mRNA transcripts that correspond to LFG4. A Northern blot containing total RNAs from various human tissues was used (Human 12-Lane MTN Blot, Clontech, Palo Alto, CA), and the

EST containing 62158_at sequence was radioactively labeled by the random primer method and used to probe the blot. The blot was hybridized in 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhart's solution, and 0.2 mg/ml herring sperm DNA at 42°C and washed with 0.2X SSC containing 0.1% SDS at room temperature. The Northern blot showed a single transcript for this gene, which is approximately 5.4 kb in size. This corresponds to the size of the LFG4 clone (SEQ ID NO: 9).

EXAMPLE 9: Identification of Differentially Expressed mRNA in Cancers - 5

The process in EXAMPLE 1 was repeated except that the marker LFG5 was used instead of the marker LFG1.

Analysis of the chip data showed that the expression of the marker LFG5 was significantly down-regulated in cancer tissue samples compared to samples from normal tissue. The expression level of LFG5 (SEQ ID NO: 11) can be measured by chip sequence fragment no. 46659_at on Affymetrix GeneChips® U95. The expression levels of 46659_at in various malignant neoplasms, compared to normal control tissues, are shown in Table 6, where the fold-change, the direction of the change (up- or down-regulation), p-value are also indicated. The fold change (cancerous/normal) was calculated by comparing the geometric mean of average difference in a cancerous sample set against the geometric mean of average difference in the normal tissue sample set. Also indicated in the Table 6 are, for each tissue type, the numbers of samples that are called present, absent, or marginal together with the total number of samples in that sample set. These data indicate that differential regulation of LFG5 may be diagnostic for cancer.

TABLE 6

Tissue	Pathology / Morphology	Geometric Mean	Number of Samples				Fold Change	Direction	p-value
			Total	Present	Marginal	Absent			
BREAST	NORMAL TISSUE, NOS	152.75	34	31	0	3			
	INFILTRATING DUCT CARCINOMA	404.58	61	60	0	1	2.65	up	0
	INFILTRATING LOBULAR CARCINOMA	277.71	10	10	0	0	1.82	up	0.07445
ESOPHAGUS	NORMAL TISSUE, NOS	85.47	18	15	0	2			
	ADENOCARCINOMA, NOS	373.97	8	8	0	0	4.38	up	0.0009
KIDNEY	NORMAL TISSUE, NOS	53.58	25	17	0	8			
	CLEAR CELL CARCINOMA	161.36	11	11	0	0	3.01	up	0.00011
	RENAL CELL CARCINMA	249.37	16	16	0	0	4.65	up	0
LUNG	NORMAL TISSUE, NOS	330.65	32	31	0	1			
	ADENOCARCINOMA, NOS	195.43	39	35	0	4	1.69	down	0.00538
LYMPH NODE	NORMAL TISSUE, NOS	219.77	9	9	0	0			
	MALIGNANT LYMPHOMA, NOS	142.09	12	11	0	1	1.55	down	0.25114
OVARY	NORMAL TISSUE, NOS	90.40	23	19	0	4			
	PAPILLARY SEROUS ADENOCARCINOMA	418.81	23	23	0	0	4.63	up	0
PANCREAS	NORMAL TISSUE, NOS	38.53	20	12	0	8			
	ADENOCARCINOMA, NOS	344.37	25	25	0	0	8.94	up	0
STOMACH	NORMAL TISSUE, NOS	185.50	18	17	0	1			
	ADENOCARCINOMA, NOS	279.62	38	35	0	3	1.51	up	0.12664

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 46659_at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer). PCR primers (5'-AAGGCTTTATCAGGTCTGCATATAGAATC-3' (SEQ ID NO: 51) and 5'-GCAAAGAACCCTAATGCTATTTATCAGC-3' (SEQ ID NO: 52)) designed based on the sequence information file of the specific Affymetrix fragment (46659_at) were used in the assay. The target gene in each RNA sample (10 ng of total RNA) was assayed relative to a reference gene. For this purpose, primers (5'-GTTTTTCCTAATTTTGGCATGAAC-3' (SEQ ID NO: 19) and 5'-CGCCCAAGCTTTTCCTTTT-3' (SEQ ID NO: 20)) specific to the CTBP1 gene (C-terminal binding protein 1) were used to serve as control primers. This approach provides the relative expression as measured by cycle threshold (Ct) value of the target mRNA relative to an amount of CTBP1 Ct value. The sample panel included total RNA pairs of normal and tumor tissues from kidney, lung, ovary, and pancreas (Ambion, Inc., Austin, TX). The Q-RT-PCR data confirms the differential regulation of LFG5 in cancer compared to normal samples.

EXAMPLE 10: Cloning of Full-Length Human cDNA (LFG5) Corresponding to Differentially Expressed mRNA Species

The full-length cDNA having SEQ ID NO: 11 was obtained by the oligo-pulling method using the GeneTrapper assay (Life Technologies, Rockville, MD). Briefly, a gene-specific oligo (5'-GAGAAGACCAGGGAAGAAGCAG-3' (SEQ ID NO: 53)) was designed based on the sequence of an EST containing 46659_at sequence. The oligo was labeled with biotin and used to hybridize with 5 μg of single strand plasmid DNA (cDNA recombinants) from a human heart library (ResGen, Huntsville, AL) following the procedures of Sambrook *et al.* The hybridized cDNAs were separated by streptavidin-conjugated beads and eluted by heating. The eluted cDNA was converted to double strand plasmid DNA and used to transform *E. coli* cells (DH10B) and the longest cDNA was

screened. After positive selection was confirmed by PCR using gene-specific primers, the cDNA clone was subjected to DNA sequencing.

The nucleotide sequence of the full-length human cDNAs corresponding to the differentially regulated mRNA detected above is set forth in SEQ ID NO: 11. The cDNA comprises 3098 base pairs.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 11, at nucleotides 223-1569 (223-1572 including the stop codon), encodes a protein of 449 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 11 is set forth in SEQ ID NO: 12.

SEQ ID NO: 12 contains a thymidylate kinase domain (amino acid positions 257-438). Thymidylate kinase is a member of nucleotide monophosphate kinases (NMPKs) which play roles in the nucleotide synthesis for RNA and DNA synthesis and are required for the pharmacological activation of therapeutic nuicleoside and nucleotide analogs (Van Rompay *et al.* (2000), *Pharmacology & Therapeutics* 87:189-198). SEQ ID NO: 12 exhibits homology to a mouse thymidylate kinase (GenBank Accession No. NM_020557) which is induced during macrophage activation (Lee and O'Brien (1995), *J Immunol.* 154:6094-6102). It shows 63% identity over the entire contiguous sequence.

Figure 7 shows the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NO: 12 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.* 157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above.

Analysis by Northern blot was performed to determine the size of the mRNA transcripts that correspond to LFG5. A Northern blot containing total RNAs from various human tissues was used (Human MTN Blot, Clontech, Palo Alto, CA), and an EST containing 82941_at sequence was radioactively labeled by the random primer method and

used to probe the blot. The blot was hybridized in 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhart's solution, and 0.2 mg/ml herring sperm DNA at 42°C and washed with 0.2X SSC containing 0.1% SDS at room temperature. The Northern blot showed a single transcript for this gene, which is approximately 3.0 kb in size. This corresponds to the size of the LFG5 clone (SEQ ID NO: 11).

EXAMPLE 11: Identification of Differentially Expressed mRNA in Cancers - 6

The process in EXAMPLE 1 was repeated except that the marker LFG6 was used instead of the marker LFG1.

Analysis of the chip data showed that the expression of the marker LFG6 was significantly up-regulated in cancer tissue samples compared to samples from normal tissue. The expression level of LFG6 (SEQ ID NO: 13 or 15) can be measured by chip sequence fragment no. 44103_at on Affymetrix GeneChips® U95. The 44103_at sequence is derived from the EST AA865614. The expression levels of 44103_at in various malignant neoplasms, compared to normal control tissues, are shown in Table 7, where the fold-change, the direction of the change (up- or down-regulation), p-value are also indicated. The fold change (cancerous/normal) was calculated by comparing the geometric mean of average difference in a cancerous sample set against the geometric mean of average difference in the normal tissue sample set. A fold change greater than 1.5 was considered to be significant (Wodicka *et al.* (1997), *Nature Biotech.* 15:1359-1367). Also indicated in the Table 7 are, for each tissue type, the numbers of samples that are called present, absent, or marginal together with the total number of samples in that sample set. These data indicate that up-regulation of LFG6 may be diagnostic for cancer.

TABLE 7

Tissue	Pathology / Morphology	Geometric Mean	Number of Samples				Fold Change	Direction	p-value
			Total	Present	Marginal	Absent			
KIDNEY	NORMAL TISSUE, NOS	337.71	25	25	0	0			
	CLEAR CELL ADENOCARCINOMA, NOS	556.82	11	11	0	0	1.65	up	0.00314
LIVER	NORMAL TISSUE, NOS	406.93	19	18	0	1			
	HEPATOCELLULAR CARCINOMA, NOS	619.40	23	22	0	1	1.52	up	0.00303
OVARY	NORMAL TISSUE, NOS	380.10	23	23	0	0			
	PAPILLARY SEROUS ADENOCARCINOMA	578.60	23	23	0	0	1.52	up	0.00013
PANCREAS	NORMAL TISSUE, NOS	138.75	20	11	1	8			
	ADENOCARCINOMA, NOS	453.01	25	25	0	0	3.26	up	0.00002

The GeneChip expression results, determined by sample binding to chip sequence fragment no. 44103_at, were validated by quantitative RT-PCR (Q-RT-PCR) using the Taqman® assay (Perkin-Elmer). PCR primers (5'-GGACGGGGAACTTGGACGC-3' (SEQ ID NO: 54) and 5'-AAGTGCAGGGCCTCTGGGTG-3' (SEQ ID NO: 55)) designed based on the sequence information file of the specific Affymetrix fragment (44103_at) were used in the assay. The target gene in each RNA sample (10 ng of total RNA) was assayed relative to a reference gene. For this purpose, primers (5'-GTTTTTCCTAATTTTGGCATGAAC-3' (SEQ ID NO: 19) and 5'-CGCCCAAGCTTTTCCTTTT-3' (SEQ ID NO: 20)) specific to the CTBP1 gene (C-terminal binding protein 1) were used to serve as control primers. This approach provides the relative expression as measured by cycle threshold (Ct) value of the target mRNA relative to an amount of CTBP1 Ct value. The sample panel included total RNA pairs of normal and tumor tissues from liver and ovary (Ambion, Inc., Austin, TX). The Q-RT-PCR data confirms the up-regulation of LFG6 in cancer compared to normal samples.

EXAMPLE 12: Cloning of Full-Length Human cDNA (LFG6) Corresponding to Differentially Expressed mRNA Species

The full-length cDNA having SEQ ID NO: 13 or 15 was obtained by the oligo-pulling method using the GeneTrapper assay (Life Technologies, Rockville, MD). Briefly, a gene-specific oligo (5'-CGCTGGGTCATCGGACGGT-3' (SEQ ID NO: 56)) was designed based on the sequence of an EST containing 44103_at sequence. The oligo was labeled with biotin and used to hybridize with 5 μg of single strand plasmid DNA (cDNA recombinants) from a fully differentiated human stomach adenocarcinoma library (ResGen, Huntsville, AL) following the procedures of Sambrook *et al.* The hybridized cDNAs were separated by streptavidin-conjugated beads and eluted by heating. The eluted cDNA was converted to double strand plasmid DNA and used to transform *E. coli* cells (DH10B) and

the longest cDNA was screened. After positive selection was confirmed by PCR using gene-specific primers, the cDNA clone was subjected to DNA sequencing.

The nucleotide sequence of the full-length human cDNAs corresponding to the differentially regulated mRNA detected above is set forth in SEQ ID NOS: 13 and 15. In the former, the cDNA comprises 1893 base pairs. In the latter, the cDNA comprises 1597 base pairs.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 13, at nucleotides 418-1392 (418-1395 including the stop codon), encodes a protein of 325 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 13 is set forth in SEQ ID NO: 14. Figure 9 shows the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NO: 14 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.* 157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above.

An open reading frame within the cDNA nucleotide sequence of SEQ ID NO: 15, at nucleotides 271-1431 (271-1434 including the stop codon), encodes a protein of 387 amino acids. The amino acid sequence corresponding to a predicted protein encoded by SEQ ID NO: 15 is set forth in SEQ ID NO: 16. Figure 10 shows the results of a hydrophobicity analysis of the amino acid sequence of SEQ ID NO: 16 using Kyte-Doolittle values (Kyte and Doolittle (1982), *J. Mol. Biol.* 157:105-142). Hydrophilic regions may be used to produce antigenic peptides, as described above.

SEQ ID NOS: 14 and 16 contain ubiquitin homologues (UBQ) domain (amino acid positions 239-300). SEQ ID NOS: 14 and 16 are similar to rat Sharpin protein (Lim *et al.* (2001), *Mol Cell Neurosci* 17:385-397). Sharpin directly interacts with the ankyrin repeats of Shank protein which functions in the organization of cytoskeletal complexes and intracellular signaling at specialized cell junctions (Sheng and Kim (2000), *J Cell Sci* 113:1851-1856).

Analysis by Northern blot was performed to determine the size of the mRNA transcripts that correspond to LFG6. A Northern blot containing total RNAs from various human tissues was used (Human 12-Lane MTN Blot, Clontech, Palo Alto, CA), and an EST containing 44103_at sequence was radioactively labeled by the random primer method and used to probe the blot. The blot was hybridized in 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhart's solution, and 0.2 mg/ml herring sperm DNA at 42°C and washed with 0.2X SSC containing 0.1% SDS at room temperature. The Northern blot showed three transcripts for this gene, which are approximately 2.2 kb, 1.5 kb, and 1.2 kb in size. This corresponds to the sizes of the LFG6 clones (SEQ ID NO: 13 and 15).

Although the present invention has been described in detail with reference to examples above, it is understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims. All cited patents, patent applications and publications referred to in this application are herein incorporated by reference in their entirety.